

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:14:14 ; Search time 16.7283 Seconds  
(without alignments)  
75.324 Million cell updates/sec

Title: US-09-719-379A-5  
Perfect score: 107  
Sequence: 1 RSDYKFYEDANGTRDHKKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS.COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 107   | 100.0       | 40     | 2  | US-08-460-502-10     |
| 2          | 107   | 100.0       | 40     | 4  | US-09-148-711A-10    |
| 3          | 107   | 100.0       | 359    | 1  | US-08-457-957B-2     |
| 4          | 107   | 100.0       | 359    | 3  | US-08-467-722A-2     |
| 5          | 107   | 100.0       | 359    | 4  | US-09-451-184-2      |
| 6          | 101   | 94.4        | 18     | 2  | US-08-460-502-1      |
| 7          | 101   | 94.4        | 18     | 4  | US-09-148-711A-1     |
| 8          | 79    | 73.8        | 338    | 1  | US-08-210-394-1      |
| 9          | 47    | 43.9        | 1561   | 3  | US-08-894-017-23     |
| 10         | 47    | 43.9        | 1561   | 4  | US-08-894-017-23     |
| 11         | 47    | 43.9        | 1565   | 6  | US-09-456-474-23     |
| 12         | 44.5  | 41.6        | 442    | 4  | US-09-328-352-5419   |
| 13         | 43    | 40.2        | 366    | 3  | US-09-718-692-4      |
| 14         | 43    | 40.2        | 366    | 4  | US-09-718-852-4      |
| 15         | 43    | 40.2        | 366    | 4  | US-09-718-815-4      |
| 16         | 43    | 40.2        | 780    | 1  | US-08-485-621-2      |
| 17         | 43    | 40.2        | 780    | 2  | US-08-973-831-2      |
| 18         | 43    | 40.2        | 780    | 5  | PCT-US96-09530A-2    |
| 19         | 43    | 40.2        | 1637   | 3  | US-09-718-692-2      |
| 20         | 43    | 40.2        | 1637   | 4  | US-09-718-852-2      |
| 21         | 43    | 40.2        | 1637   | 4  | US-09-718-815-2      |
| 22         | 42    | 39.3        | 189    | 4  | US-09-198-452A-135   |
| 23         | 42    | 39.3        | 449    | 4  | US-09-248-796A-16049 |
| 24         | 42    | 39.3        | 693    | 4  | US-09-081-385-154    |
| 25         | 41    | 38.3        | 623    | 4  | US-09-248-796A-14256 |
| 26         | 41    | 38.3        | 744    | 2  | US-08-462-080B-2     |
| 27         | 41    | 38.3        | 744    | 3  | US-08-462-090-2      |

|    |      |      |      |   |                      |                   |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 41   | 38.3 | 744  | 3 | US-08-463-461-2      | Sequence 2, Appli |
| 29 | 41   | 38.3 | 756  | 4 | US-09-248-796A-17753 | Sequence 17753, A |
| 30 | 40.5 | 37.9 | 538  | 2 | US-08-867-941-18     | Sequence 18, Appl |
| 31 | 40.5 | 37.9 | 538  | 3 | US-09-074-658-18     | Sequence 18, Appl |
| 32 | 40.5 | 37.9 | 541  | 2 | US-08-867-941-14     | Sequence 14, Appl |
| 33 | 40.5 | 37.9 | 541  | 3 | US-09-074-658-14     | Sequence 14, Appl |
| 34 | 40.5 | 37.9 | 544  | 4 | US-09-540-236-2760   | Sequence 2760, Ap |
| 35 | 40.5 | 37.9 | 2432 | 3 | US-09-074-658-15     | Sequence 15, Appl |
| 36 | 40.5 | 37.9 | 2439 | 3 | US-09-074-658-11     | Sequence 11, Appl |
| 37 | 40   | 37.4 | 256  | 3 | US-09-134-001C-5471  | Sequence 5471, Ap |
| 38 | 40   | 37.4 | 322  | 3 | US-09-080-855-13     | Sequence 13, Appl |
| 39 | 40   | 37.4 | 322  | 4 | US-09-566-076-13     | Sequence 13, Appl |
| 40 | 40   | 37.4 | 336  | 4 | US-09-248-796A-18799 | Sequence 18799, A |
| 41 | 40   | 37.4 | 340  | 3 | US-09-134-001C-4056  | Sequence 4056, Ap |
| 42 | 40   | 37.4 | 511  | 4 | US-09-248-796A-16380 | Sequence 16380, A |
| 43 | 39   | 36.4 | 91   | 4 | US-09-270-767-49094  | Sequence 49094, A |
| 44 | 39   | 36.4 | 194  | 4 | US-09-107-532A-6229  | Sequence 6229, Ap |
| 45 | 39   | 36.4 | 204  | 4 | US-09-252-991A-26646 | Sequence 26646, A |

ALIGNMENTS

RESULT 1  
US-08-460-502-10  
; Sequence 10, Application US/08460502  
; Patent No. 5843464  
; GENERAL INFORMATION:  
; APPLICANT: Bakaletz, Lauren O.  
; APPLICANT: Kaumaya, Parvin T.  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDING ADDRESSES:  
; ADDRESSEE: Calfee, Halter and Griswold  
; STREET: 800 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2688  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,502  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goltick, Mary E.  
; REGISTRATION NUMBER: 34,829  
; REFERENCE/DOCKET NUMBER: 22727/00120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 622-8458  
; TELEFAX: (216) 241-0816  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-460-502-10

Query Match 100.0%; Score 107; DB 2; Length 40;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19  
Db 1 RSDYKFYEDANGTRDHKKG 19

```

Db          138 RSDYKFYEDANGTRDHKKG 156
|||||
RESULT 4
US-08-467-722A-2
; Sequence 2, Application US/08467722A
; Patent No. 6030626
; GENERAL INFORMATION:
; APPLICANT: Kolattukudy, P. E.
; TITLE OF INVENTION: Otitis Media Vaccine
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter and Griswold
; STREET: Suite 1800 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A>
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,722A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 22727/00102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8458
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-467-722A-2
Query Match 100.0%; Score 107; DB 3; Length 359;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels

QY          1 RSDYKFYEDANGTRDHKKG 19
|||||
Db          138 RSDYKFYEDANGTRDHKKG 156
|||||
RESULT 5
US-09-451-184-2
; Sequence 2, Application US/09451184
; Patent No. 6562349
; GENERAL INFORMATION:
; APPLICANT: Kolattukudy, P. E.
; TITLE OF INVENTION: Otitis Media Vaccine
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter and Griswold
; STREET: 1400 McDonald Investment Center,
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A>
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/451,184  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Docherty, Pamela A.  
;; REGISTRATION NUMBER: 40,591  
;; REFERENCE/DOCKET NUMBER: 24547/04000  
;; TELEPHONE: (216) 622-8416  
;; TELEFAX: (216) 241-0816  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 359 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-451-184-2

Query Match 100.0%; Score 107; DB 4; Length 359;  
Best Local Similarity 100.0%; Pred. No. 2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKK 19  
Db 138 RSDYKFYEDANGTRDHKK 156  
|||||

RESULT 6  
US-08-460-502-1  
; Sequence 1, Application US/08460502  
; Patent No. 5843464  
; GENERAL INFORMATION:  
; APPLICANT: Bakaletz, Lauren O.  
; APPLICANT: Kaumaya, Parvin T.  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calfee, Halter and Griswold  
; STREET: 800 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2688  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Golrick, Mary E.  
; REGISTRATION NUMBER: 34,829  
; REFERENCE/DOCKET NUMBER: 22727/00120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 622-8458  
; TELEFAX: (216) 241-0816  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-460-502-1

Query Match 94.4%; Score 101; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 6.4e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKK 18  
Db 1 RSDYKFYEDANGTRDHKK 18  
|||||

RESULT 7  
US-09-148-711A-1  
; Sequence 1, Application US/09148711A  
; Patent No. 6436405  
; GENERAL INFORMATION:  
; APPLICANT: The Ohio State University  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; FILE REFERENCE: 18525-04010  
; CURRENT APPLICATION NUMBER: US/09/148,711A  
; CURRENT FILING DATE: 1998-09-04  
; PRIOR APPLICATION NUMBER: 08/460,502  
; PRIOR FILING DATE: 1995-06-02  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Haemophilus influenza  
US-09-148-711A-1

Query Match 94.4%; Score 101; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 6.4e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKK 18  
Db 1 RSDYKFYEDANGTRDHKK 18  
|||||

RESULT 8  
US-08-210-394-1  
; Sequence 1, Application US/08210394  
; Patent No. 5770213  
; GENERAL INFORMATION:  
; APPLICANT: Zlotnick Dr., Gary W.  
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus  
; TITLE OF INVENTION: Influenzae P5 Protein as a Vaccine  
; TITLE OF INVENTION: Haemophilus Influenzae Strain  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: US  
; ZIP: 07470-8426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/210,394  
; FILING DATE: 07-MAR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J  
; REFERENCE/DOCKET NUMBER: 32,144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201/831-3246  
; TELEFAX: 201/831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein





```
Db 36 FVEDANGSILPRTGYLTRDKKQG 58

RESULT 13
US-09-718-692-4
; Sequence 4, Application US/09718692
; Patent No. 6383796
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6383796el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718,692
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Human
US-09-718-692-4

Query Match 40.2%; Score 43; DB 3; Length 366;
Best Local Similarity 66.7%; Pred. No. 49;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANG 12
||:|:|||||
Db 166 RSNIKIHEDANG 177

RESULT 14
US-09-718-852-4
; Sequence 4, Application US/09718852
; Patent No. 6426193
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6426193el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718,852
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Human
US-09-718-852-4

Query Match 40.2%; Score 43; DB 4; Length 366;
Best Local Similarity 66.7%; Pred. No. 49;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANG 12
||:|:|||||
Db 166 RSNIKIHEDANG 177

RESULT 15
US-09-718-815-4
; Sequence 4, Application US/09718815
; Patent No. 6455293
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6455293el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718,815
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Human
US-09-718-815-4

Query Match 40.2%; Score 43; DB 4; Length 366;
Best Local Similarity 66.7%; Pred. No. 49;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANG 12
||:|:|||||
Db 166 RSNIKIHEDANG 177

RESULT 16
US-08-485-621-2
; Sequence 2, Application US/08485621
; Patent No. 5691187
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5691187is
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,621
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TPU-1584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-621-2

Query Match 40.2%; Score 43; DB 1; Length 780;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANGTRDH 16
||:|:|||||
Db 382 QSDPKFKFETAKLRDH 397

RESULT 17
US-08-973-831-2
; Sequence 2, Application US/08973831
; Patent No. 5942386
```

```
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5942386ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,831
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,621
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-973-831-2

Query Match 40.2%; Score 43; DB 2; Length 780;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDH 16
   :|||:|:|:|
Db 382 QSDFKKFTARKLRDH 397

RESULT 18
PCT-US96-09530A-2
; Sequence 2, Application PC/TUS9609530A
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09530A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,621
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,399
; FILING DATE: 21-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1970
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-09530A-2

Query Match 40.2%; Score 43; DB 5; Length 780;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDH 16
   :|||:|:|:|
Db 382 QSDFKKFTARKLRDH 397

RESULT 19
US-09-718-692-2
; Sequence 2, Application US/09718692
; Patent No. 6383796
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6383796el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718,692
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1637
; TYPE: PRT
; ORGANISM: Human
; US-09-718-692-2

Query Match 40.2%; Score 43; DB 3; Length 1637;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANG 12
   ||:|:||||
Db 171 RSNIKIHEDANG 182

RESULT 20
US-09-718-852-2
; Sequence 2, Application US/09718852
; Patent No. 6426193
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6426193el motor proteins and methods for
```

;; TITLE OF INVENTION: their use  
;; FILE REFERENCE: 1052  
;; CURRENT APPLICATION NUMBER: US/09/718.852  
;; CURRENT FILING DATE: 2000-11-22  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSEQ for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 1637  
;; TYPE: PRT  
;; ORGANISM: Human  
US-09-718-852-2

Query Match 40.2%; Score 43; DB 4; Length 1637;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANG 12  
||:|:|||||  
Db 171 RSNIKIHEDANG 182

## RESULT 21

US-09-718-815-2  
; Sequence 2, Application US/09718815  
; Patent No. 6455293

;; GENERAL INFORMATION:  
;; APPLICANT: Beraud, Christophe  
;; APPLICANT: Freedman, Richard  
;; TITLE OF INVENTION: No. 6455293el motor proteins and methods for  
;; TITLE OF INVENTION: their use

;; FILE REFERENCE: 1052  
;; CURRENT APPLICATION NUMBER: US/09/718.815  
;; CURRENT FILING DATE: 2000-11-22  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSEQ for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 1637

;; TYPE: PRT  
;; ORGANISM: Human  
US-09-718-815-2

Query Match 40.2%; Score 43; DB 4; Length 1637;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANG 12  
||:|:|||||  
Db 171 RSNIKIHEDANG 182

## RESULT 22

US-09-198-452A-135  
; Sequence 135, Application US/09198452A  
; Patent No. 6559294

;; GENERAL INFORMATION:  
;; APPLICANT: Griffois, R.  
;; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
;; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
;; TITLE OF INVENTION: and treatment of infection  
;; FILE REFERENCE: 9710-003-999  
;; CURRENT APPLICATION NUMBER: US/09/198.452A  
;; CURRENT FILING DATE: 1998-11-24  
;; NUMBER OF SEQ ID NOS: 6849

;; SEQ ID NO 135  
;; LENGTH: 189  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-135

Query Match 39.3%; Score 42; DB 4; Length 189;  
Best Local Similarity 50.0%; Pred. No. 34;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKFEDANGTROHKK 18  
||:|:|||||  
Db 58 DHKFEDETTNRDHF 73

## RESULT 23

US-09-248-796A-16049  
; Sequence 16049, Application US/09248796A  
; Patent No. 6747137

;; GENERAL INFORMATION:  
;; APPLICANT: Keith Weinstock et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 107196.132  
;; CURRENT APPLICATION NUMBER: US/09/248.796A  
;; CURRENT FILING DATE: 1999-02-12  
;; PRIOR APPLICATION NUMBER: US 60/074,725  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 16049  
;; LENGTH: 449

;; TYPE: PRT  
;; ORGANISM: Candida albicans  
;; FEATURE:  
;; NAME/KEY: UNSURE  
;; LOCATION: (433), (444)  
;; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown

US-09-248-796A-16049

Query Match 39.3%; Score 42; DB 4; Length 449;  
Best Local Similarity 58.3%; Pred. No. 89;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKFVEDANGT 13  
||:|:|||||  
Db 369 SDFNFYQDLNPT 380

## RESULT 24

US-09-081-385-154  
; Sequence 154, Application US/09081385  
; Patent No. 6593456

;; GENERAL INFORMATION:  
;; APPLICANT: Gatanaga, T.  
;; APPLICANT: Granger, G.A.  
;; TITLE OF INVENTION: Factors Altering Tumor Necrosis  
;; TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods  
;; TITLE OF INVENTION: of Use Thereof  
;; NUMBER OF SEQUENCES: 154

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MORRISON & FOERSTER  
;; STREET: 755 PAGE MILL ROAD  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: Windows  
;; SOFTWARE: FastSEQ for Windows Version 2.0b  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/081.385  
;; FILING DATE:

;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/964,747  
;; FILING DATE: 05-NOV-1997  
;; APPLICATION NUMBER: 60/030,761  
;; FILING DATE: 06-NOV-1996  
;; ATTORNEY/AGENT INFORMATION:

```

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/462,080B
/ FILING DATE: 05-JUN-1995
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/248,586
/ FILING DATE: 24-MAY-1994
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/807,028
/ FILING DATE: 10-DEC-1991
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/625,140
/ FILING DATE: 10-DEC-1990
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stone, Christopher L.
/ REGISTRATION NUMBER: 35,696
/ REFERENCE/DOCKET NUMBER: GC78D3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-846-7555
/ TELEFAX: 650-845-6504
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 744 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-462-080B-2

Query Match 38.3%; Score 41; DB 2; Length 744;
Best Local Similarity 47.1%; Pred. No. 2.3e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps

QY 3 DYKFVEDANGTRDHKKG 19
    ||| :||| :|||
Db 589 DYKHFDDANITPYEFG 605

RESULT 27
US-08-462-090-2
/ Sequence 2, Application US/08462090
/ Patent No. 6022725
/ GENERAL INFORMATION:
/ APPLICANT: Fowler, Timothy
/ APPLICANT: Barnett, Christopher C.
/ APPLICANT: Shoemaker, Sharon
/ TITLE OF INVENTION: Saccharification of Cellulose by Cloning
/ TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
/ TITLE OF INVENTION: Trichoderma Reesei
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Burns, Doane, Swecker & Mathis
/ STREET: George Mason Building, 699 Prince St.
/ CITY: Alexandria
/ STATE: Virginia
/ COUNTRY: U.S.A.
/ ZIP: 22313-1404
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/462,090
/ FILING DATE: 05-JUN-1995
/ CLASSIFICATION: 435

```

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/625,140  
;; FILING DATE: 10-DEC-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dillahunty, T. Gene  
;; REGISTRATION NUMBER: 25,423  
;; REFERENCE/DOCKET NUMBER: 010055-056  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-854-7400  
;; TELEFAX: 415-854-8275  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 744 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-462-090-2

Query Match 38.3%; Score 41; DB 3; Length 744;  
Best Local Similarity 47.1%; Pred. No. 2.3e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DYKFYEDANGTRDHKKG 19  
||| :||| : :  
Db 589 DYKHFDANITPRYFEG 605

## RESULT 28

US-08-463-461-2  
; Sequence 2, Application US/08463461  
; Patent No. 6103464

;; GENERAL INFORMATION:  
;; APPLICANT: Fowler, Timothy  
;; APPLICANT: Barnett, Christopher C.  
;; APPLICANT: Shoenaker, Sharon  
;; TITLE OF INVENTION: Saccharification of Cellulose by Cloning  
;; TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of  
;; TITLE OF INVENTION: Trichoderma Reesei  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genencor International, Inc.  
;; STREET: 925 Page Mill Road  
;; CITY: Palo Alto  
;; STATE: California  
;; COUNTRY: U.S.A.  
;; ZIP: 94304-1013

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/463,461  
;; FILING DATE: 05-JUN-1995  
;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Christopher L. Stone  
;; REGISTRATION NUMBER: 35,696  
;; REFERENCE/DOCKET NUMBER: GC78D4  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-846-7555  
;; TELEFAX: 415-845-6504  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 744 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-463-461-2

Query Match 38.3%; Score 41; DB 3; Length 744;  
Best Local Similarity 47.1%; Pred. No. 2.3e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DYKFYEDANGTRDHKKG 19  
||| :||| : :  
Db 589 DYKHFDANITPRYFEG 605

## RESULT 29

US-09-248-796A-17753  
; Sequence 17753, Application US/09248796A  
; Patent No. 6747137

;; GENERAL INFORMATION:  
;; APPLICANT: Keith Weinstock et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.132  
;; CURRENT APPLICATION NUMBER: US/09/248,796A  
;; CURRENT FILING DATE: 1999-02-12  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/074,725  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 17753  
;; LENGTH: 756  
;; TYPE: PRT  
;; ORGANISM: Candida albicans  
;; FEATURE:  
;; NAME/KEY: UNSURE  
;; LOCATION: (21)  
;; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno

US-09-248-796A-17753

Query Match 38.3%; Score 41; DB 4; Length 756;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANGTRDHK 17  
||| :||| : :  
Db 532 SDLKFTSLNGVNSDK 547

## RESULT 30

US-08-867-941-18  
; Sequence 18, Application US/08867941  
; Patent No. 5977337

;; GENERAL INFORMATION:  
;; APPLICANT: Loosmore, Sheena M  
;; APPLICANT: Du, Run-Pan  
;; APPLICANT: Wang, Quijun  
;; APPLICANT: Yang, Yan-Ping  
;; APPLICANT: Klein, Michel H  
;; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
;; NUMBER OF SEQUENCES: 67  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sim & McBurney  
;; STREET: 6th Floor, 330 University Avenue  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: M5G 1R7

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/867,941  
;; FILING DATE: 03-JUN-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-867-941-18

Query Match 37.9%; Score 40.5; DB 2; Length 538;  
Best Local Similarity 35.7%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| | | | | : | : | : |  
Db 310 DGKFYDDKSHNELAVFAHAGLRKDHQKG 337

RESULT 31  
US-09-074-658-18  
Sequence 18, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-18

Query Match 37.9%; Score 40.5; DB 3; Length 538;  
Best Local Similarity 35.7%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| | | | | : | : | : |  
Db 310 DGKFYDDKSHNELAVFAHAGLRKDHQKG 337

RESULT 32

US-08-867-941-14  
Sequence 14, Application US/08867941  
Patent No. 5977337  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Du, Run-Pan  
APPLICANT: Wang, Quijun  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/867,941  
FILING DATE: 03-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jbb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 541 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-867-941-14

Query Match 37.9%; Score 40.5; DB 2; Length 541;  
Best Local Similarity 35.7%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| | | | | : | : | : |  
Db 313 DGKFYDDKSHNELAVFAHAGLRKDHQKG 340

RESULT 33  
US-09-074-658-14  
Sequence 14, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/074,658  
;; FILING DATE: 08-MAY-1998  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-795  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 541 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-09-074-658-14

Query Match 37.9%; Score 40.5; DB 3; Length 541;  
Best Local Similarity 35.7%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| |||:| | :|||  
Db 313 DGKFYDDKSHNELAVFAHAGLRKDHQKG 340

## RESULT 14

US-09-540-236-2760  
; Sequence 2760, Application US/09540236  
; Patent No. 6673910

## GENERAL INFORMATION:

;; APPLICANT: Gary L. Breton et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATABOLISM  
;; FILE REFERENCE: 2709.2005-001  
;; CURRENT APPLICATION NUMBER: US/09/540,236  
;; CURRENT FILING DATE: 2000-04-04

;; NUMBER OF SEQ ID NOS: 3840  
;; SEQ ID NO 2760  
;; LENGTH: 544  
;; TYPE: PRT

;; ORGANISM: M.catarrhalis

US-09-540-236-2760

Query Match 37.9%; Score 40.5; DB 4; Length 544;  
Best Local Similarity 35.7%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| |||:| | :|||  
Db 316 DGKFYDDKSHNELAVFAHAGLRKDHQKG 343

## RESULT 35

US-09-074-658-15  
; Sequence 15, Application US/09074658  
; Patent No. 6184371

## GENERAL INFORMATION:

;; APPLICANT: Loosmore, Sheena M  
;; APPLICANT: Run-Pan Du  
;; APPLICANT: Quijun Wang  
;; APPLICANT: Yang, Yan-Ping  
;; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
;; NUMBER OF SEQUENCES: 78  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sim & McBurney  
;; STREET: 6th Floor, 330 University Avenue  
;; CITY: Toronto

;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: MSG 1R7  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/074,658  
;; FILING DATE: 08-MAY-1998  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-795  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2432 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-09-074-658-15

Query Match 37.9%; Score 40.5; DB 3; Length 2432;  
Best Local Similarity 35.7%; Pred. No. 1e+03;  
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy 3 DYKFYED-----ANGTRDHKKG 19  
| |||:| | :|||  
Db 2204 DGKFYDDKSHNELAVFAHAGLRKDHQKG 2231

## RESULT 36

US-09-074-658-11  
; Sequence 11, Application US/09074658  
; Patent No. 6184371

## GENERAL INFORMATION:

;; APPLICANT: Loosmore, Sheena M  
;; APPLICANT: Run-Pan Du  
;; APPLICANT: Quijun Wang  
;; APPLICANT: Yang, Yan-Ping  
;; APPLICANT: Klein, Michel H  
;; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
;; NUMBER OF SEQUENCES: 78  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sim & McBurney  
;; STREET: 6th Floor, 330 University Avenue  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: MSG 1R7  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA: US/09/074,658  
;; APPLICATION NUMBER: US/09/074,658  
;; FILING DATE: 08-MAY-1998  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-795  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 2439 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-658-11

Query Match      37.9%; Score 40.5; DB 3; Length 2439;
Best Local Similarity 35.7%; Pred. No. 1e+03;
Matches 10; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

Qy      3 DYKFVED-----ANGTRDHKKG 19
Db      2211 DQKFYDDKSHNELAVFAHAGLRKHQKG 2238

RESULT 37
US-09-134-001C-5471
; Sequence 5471, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5471
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5471

Query Match      37.4%; Score 40; DB 3; Length 256;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      4 YKFYEDANGTRDH 16
Db      212 YDYVEDANTVNVH 224

RESULT 38
US-09-080-855-13
; Sequence 13, Application US/09080855A
; Patent No. 6083721
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franzn, Petra
; APPLICANT: Aspenstrm, Pontus
; APPLICANT: Hellman, Ulf
; APPLICANT: Gonez, Leonel Jorge
; APPLICANT: Heldin, Carl-Henrik
; TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
; FILE REFERENCE: L0461/7030
; CURRENT APPLICATION NUMBER: US/09/080,855A
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: 08/805,583
; EARLIER FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-080-855-13

Query Match      37.4%; Score 40; DB 3; Length 322;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
```

```
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy      6 FYEDANGTRDHKKG 19
Db      250 FHLSNATRDYQPG 263

RESULT 39
US-09-566-076-13
; Sequence 13, Application US/09566076
; Patent No. 6475775
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franzn, Petra
; APPLICANT: Aspenstrm, Pontus
; APPLICANT: Hellman, Ulf
; APPLICANT: Gonez, Leonel Jorge
; APPLICANT: Heldin, Carl-Henrik
; TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
; FILE REFERENCE: L0461/7030
; CURRENT APPLICATION NUMBER: US/09/566,076
; CURRENT FILING DATE:
; EARLIER APPLICATION NUMBER: 09/080,855
; EARLIER FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-566-076-13

Query Match      37.4%; Score 40; DB 4; Length 322;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy      6 FYEDANGTRDHKKG 19
Db      250 FHLSNATRDYQPG 263

RESULT 40
US-09-248-796A-18799
; Sequence 18799, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18799
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18799

Query Match      37.4%; Score 40; DB 4; Length 336;
Best Local Similarity 52.6%; Pred. No. 1.4e+02;
Matches 10; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

Qy      1 RSDYKFYEDA--NGTRDHK 17
Db      190 RSDYKQYHDTTNGVSTDK 208

Search completed: November 24, 2004, 09:31:42
```



Job time : 17.7283 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:29:11 ; Search time 49.0326 Seconds  
(without alignments)  
94.040 Million cell updates/sec

Title: US-09-719-379A-3

Perfect score: 72

Sequence: 1 RSDYKFDNKRID 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description         |
|------------|-------|-------------|--------|----------------------|---------------------|
| 1          | 72    | 100.0       | 13     | US-10-203-942-3      | Sequence 3, Appli   |
| 2          | 72    | 100.0       | 22     | US-10-203-942-7      | Sequence 7, Appli   |
| 3          | 43    | 59.7        | 18     | US-10-223-711-1      | Sequence 1, Appli   |
| 4          | 43    | 59.7        | 19     | US-10-203-942-1      | Sequence 5, Appli   |
| 5          | 43    | 59.7        | 28     | US-10-203-942-5      | Sequence 10, Appli  |
| 6          | 43    | 59.7        | 40     | US-10-223-711-10     | Sequence 50359, A   |
| 7          | 43    | 59.7        | 696    | US-10-282-122A-50359 | Sequence 53280, A   |
| 8          | 42    | 58.3        | 514    | US-10-282-122A-53280 | Sequence 145844, A  |
| 9          | 41    | 56.9        | 84     | US-10-424-599-145844 | Sequence 1044, Ap   |
| 10         | 40    | 55.6        | 46     | US-10-083-357-1044   | Sequence 171824, A  |
| 11         | 40    | 55.6        | 216    | US-10-424-599-171824 | Sequence 199530, A  |
| 12         | 40    | 55.6        | 418    | US-10-425-115-199530 | Sequence 440, Appli |
| 13         | 40    | 55.6        | 437    | US-10-287-274-440    |                     |

|    |    |      |      |    |                      |                    |
|----|----|------|------|----|----------------------|--------------------|
| 14 | 40 | 55.6 | 437  | 15 | US-10-282-122A-42714 | Sequence 42714, A  |
| 15 | 40 | 55.6 | 437  | 15 | US-10-282-122A-74928 | Sequence 74928, A  |
| 16 | 40 | 55.6 | 484  | 15 | US-10-282-122A-73481 | Sequence 73481, A  |
| 17 | 39 | 54.2 | 70   | 17 | US-10-425-115-254882 | Sequence 254882, A |
| 18 | 39 | 54.2 | 115  | 18 | US-10-767-701-45949  | Sequence 45949, A  |
| 19 | 39 | 54.2 | 115  | 17 | US-10-425-115-307322 | Sequence 307322, A |
| 20 | 39 | 54.2 | 115  | 17 | US-10-425-115-307333 | Sequence 307333, A |
| 21 | 39 | 54.2 | 121  | 15 | US-10-425-114-47855  | Sequence 47855, A  |
| 22 | 39 | 54.2 | 131  | 15 | US-10-425-114-61683  | Sequence 61683, A  |
| 23 | 39 | 54.2 | 133  | 15 | US-10-425-114-42709  | Sequence 42709, A  |
| 24 | 39 | 54.2 | 136  | 15 | US-10-425-114-62912  | Sequence 62912, A  |
| 25 | 39 | 54.2 | 138  | 15 | US-10-425-114-72110  | Sequence 72110, A  |
| 26 | 39 | 54.2 | 146  | 15 | US-10-425-114-48114  | Sequence 48114, A  |
| 27 | 39 | 54.2 | 369  | 9  | US-09-893-737-50     | Sequence 50, Appli |
| 28 | 39 | 54.2 | 636  | 16 | US-10-437-963-104111 | Sequence 104111, A |
| 29 | 39 | 54.2 | 872  | 15 | US-10-282-122A-74629 | Sequence 74629, A  |
| 30 | 39 | 54.2 | 921  | 15 | US-10-282-122A-78226 | Sequence 78226, A  |
| 31 | 38 | 52.8 | 187  | 16 | US-10-437-963-141960 | Sequence 141960, A |
| 32 | 38 | 52.8 | 207  | 15 | US-10-282-122A-51668 | Sequence 51668, A  |
| 33 | 38 | 52.8 | 212  | 15 | US-10-424-599-248762 | Sequence 248762, A |
| 34 | 38 | 52.8 | 309  | 16 | US-10-437-963-193680 | Sequence 193680, A |
| 35 | 38 | 52.8 | 715  | 14 | US-10-369-493-23664  | Sequence 23664, A  |
| 36 | 38 | 52.8 | 722  | 9  | US-09-853-533A-10    | Sequence 10, Appli |
| 37 | 38 | 52.8 | 722  | 14 | US-10-360-899-10     | Sequence 10, Appli |
| 38 | 38 | 52.8 | 722  | 17 | US-10-782-141-19     | Sequence 19, Appli |
| 39 | 38 | 52.8 | 788  | 9  | US-09-908-193-35     | Sequence 35, Appli |
| 40 | 38 | 52.8 | 790  | 14 | US-10-369-493-3803   | Sequence 3803, Ap  |
| 41 | 38 | 52.8 | 843  | 10 | US-09-978-303-25     | Sequence 25, Appli |
| 42 | 38 | 52.8 | 1083 | 14 | US-10-369-493-21920  | Sequence 21920, A  |
| 43 | 38 | 52.8 | 1139 | 9  | US-09-801-574-34     | Sequence 34, Appli |
| 44 | 38 | 52.8 | 1196 | 15 | US-10-648-512-4      | Sequence 4, Appli  |
| 45 | 38 | 52.8 | 2630 | 16 | US-10-437-963-172374 | Sequence 172374, A |

#### ALIGNMENTS

RESULT 1  
US-10-203-942-3  
; Sequence 3, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOEL, PHILIPPE  
; APPLICANT: POOLMAN, JAN  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; FILE REFERENCE: B45210  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-3

Query Match 100.0%; Score 72; DB 14; Length 13;  
Best Local Similarity 100.0%; Pred. No. 5.1e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKRID 13  
| | | | | | | | | | | | |  
Db 1 RSDYKFDNKRID 13

RESULT 2

```

US-10-203-942-7
; Sequence 7, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-7

Query Match      100.0%; Score 72; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKRID 13
Db 3 RSDYKFDNKRID 15

RESULT 3
US-10-223-711-1
; Sequence 1, Application US/10223711
; Publication No. US20030113344A1
; GENERAL INFORMATION:
; APPLICANT: BAKALETZ, LAUREN T.P.
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides
; FILE REFERENCE: 18525/04058
; CURRENT APPLICATION NUMBER: US/10/223,711
; CURRENT FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: 09/148,711
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 08/460,502
; PRIOR FILING DATE: 1995-06-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-223-711-1

Query Match      59.7%; Score 43; DB 14; Length 18;
Best Local Similarity 77.8%; Pred. No. 3.5;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDN 9
Db 1 RSDYKFYED 9

RESULT 4
US-10-203-942-1
; Sequence 1, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE

```

```

; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-1

Query Match      59.7%; Score 43; DB 14; Length 19;
Best Local Similarity 77.8%; Pred. No. 3.7;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDN 9
Db 1 RSDYKFYED 9

RESULT 5
US-10-203-942-5
; Sequence 5, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-5

Query Match      59.7%; Score 43; DB 14; Length 28;
Best Local Similarity 77.8%; Pred. No. 5.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDN 9
Db 3 RSDYKFYED 11

RESULT 6
US-10-223-711-10
; Sequence 10, Application US/10223711
; Publication No. US20030113344A1
; GENERAL INFORMATION:
; APPLICANT: BAKALETZ, LAUREN O.
; APPLICANT: KAUMAYA, PRAVIN T.P.
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides
; FILE REFERENCE: 18525/04058
; CURRENT APPLICATION NUMBER: US/10/223,711

```

; CURRENT FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: 09/148,711  
; PRIOR FILING DATE: 1998-09-04  
; PRIOR APPLICATION NUMBER: 08/460,502  
; PRIOR FILING DATE: 1995-06-02  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: synthetic construct  
US-10-223-711-10

Query Match 59.7%; Score 43; DB 14; Length 40;  
Best Local Similarity 77.8%; Pred. No. 7.8;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKPYDN 9  
|||||: :  
Db 1 RSDYKFVED 9

RESULT 7  
US-10-282-122A-50359  
; Sequence 50359, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haseelbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 50359  
; LENGTH: 696  
; TYPE: PRT  
; ORGANISM: Burkholderia mallei  
; FEATURE:

; NAME/KEY: MISC FEATURE  
; LOCATION: (617)..(617)  
; OTHER INFORMATION: X-any amino acid  
US-10-282-122A-50359

Query Match 59.7%; Score 43; DB 15; Length 696;  
Best Local Similarity 46.2%; Pred. No. 1.4e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKPYDNKRID 13  
|||||: : : :  
Db 133 RSDYRIFQNRSD 145

RESULT 8  
US-10-282-122A-53280  
; Sequence 53280, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haseelbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 53280  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-10-282-122A-53280

Query Match 58.3%; Score 42; DB 15; Length 514;  
Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKPYDNK 10  
: : : : :  
Db 240 KSSYAFYDNK 249

```
RESULT 9
US-10-424-599-145844
; Sequence 145844, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 145844
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_102716C.1.pep
US-10-424-599-145844

Query Match      56.9%; Score 41; DB 15; Length 84;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 RSDYKFYDYN 9
Db      39 KSDYFFYDYN 47

RESULT 10
US-10-083-357-1044
; Sequence 1044, Application US/10083357
; Publication No. US20030054370A1
; GENERAL INFORMATION:
; APPLICANT: Qiangdong Zeng et al.
; TITLE OF INVENTION: Systemic Discovery of New Genes
; FILE REFERENCE: 032796-090
; CURRENT APPLICATION NUMBER: US/10/083,357
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 1346
; SEQ ID NO 1044
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1044

Query Match      55.6%; Score 40; DB 14; Length 46;
Best Local Similarity 58.3%; Pred. No. 27;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      1 RSDYKFYDKNRI 12
Db      17 RSSYQFYEYVKRM 28

RESULT 11
US-10-424-599-171824
; Sequence 171824, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

```
; SEQ ID NO 171824
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_126171C.1.pep
US-10-424-599-171824

Query Match      55.6%; Score 40; DB 15; Length 216;
Best Local Similarity 63.6%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 RSDYKFYDKNR 11
Db      176 KDDYPGYDKNR 186

RESULT 12
US-10-425-115-199530
; Sequence 199530, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 199530
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(418)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_113545C.1.pep
US-10-425-115-199530

Query Match      55.6%; Score 40; DB 17; Length 418;
Best Local Similarity 54.5%; Pred. No. 2.5e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy      2 SDYKFYDKNRI 12
Db      185 TNYKYTNKRM 195

RESULT 13
US-10-287-274-440
; Sequence 440, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forgyth, R. Allyn
; APPLICANT: Olesen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREO
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 440
; LENGTH: 437
```

; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-287-274-440

Query Match 55.6%; Score 40; DB 14; Length 437;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:||||  
Db 367 DYKYDN 373

## RESULT 14

US-10-282-122A-42714  
; Sequence 42714, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

## FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 42714

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-282-122A-42714

Query Match 55.6%; Score 40; DB 15; Length 437;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:||||  
Db 367 DYKYDN 373

## RESULT 15

US-10-282-122A-74928  
; Sequence 74928, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

## FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 74928

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Salmonella typhimurium

US-10-282-122A-74928

Query Match 55.6%; Score 40; DB 15; Length 437;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:||||  
Db 367 DYKYDN 373

## RESULT 16

US-10-282-122A-73481  
; Sequence 73481, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert

```
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73481
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (119)..(119)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-73481

Query Match          55.6%; Score 40; DB 15; Length 484;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9
Db 414 DYKYDN 420

RESULT 17
US-10-425-115-254882
; Sequence 254882, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 254882
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_16402C.1.pep
US-10-425-115-254882

Query Match          54.2%; Score 39; DB 17; Length 70;
Best Local Similarity 60.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFYDNKRI 12
Db 2 DYEFYEERI 11

RESULT 18
US-10-767-701-45949
; Sequence 45949, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45949
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C12309_3.pep
US-10-767-701-45949

Query Match          54.2%; Score 39; DB 16; Length 115;
Best Local Similarity 77.8%; Pred. No. 99;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 99 KFYDSKRSD 107

RESULT 19
US-10-425-115-307322
; Sequence 307322, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 307322
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_4334C.1.pep
US-10-425-115-307322

Query Match          54.2%; Score 39; DB 17; Length 115;
Best Local Similarity 77.8%; Pred. No. 99;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 99 KFYDSKRSD 107

RESULT 20
US-10-425-115-307333
; Sequence 307333, Application US/10425115
```



```

; Publication No. US20040214272A1
; GENERAL INFORMATION: Nucleic Acid Molecules and Other Molecules Associated With
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 30733
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_4335C.1.pep
US-10-425-115-30733

Query Match          54.2%; Score 39; DB 17; Length 115;
Best Local Similarity 77.8%; Pred. No. 99;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 99 KFYDSKRSD 107

RESULT 21
US-10-425-114-47855
; Sequence 47855, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 47855
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-011-H4_FLI.pep
US-10-425-114-47855

Query Match          54.2%; Score 39; DB 15; Length 121;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 105 KFYDSKRSD 113

RESULT 22
US-10-425-114-61683
; Sequence 61683, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61683
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3116-028-C4_FLI.pep
US-10-425-114-61683

Query Match          54.2%; Score 39; DB 15; Length 131;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 115 KFYDSKRSD 123

RESULT 23
US-10-425-114-42709
; Sequence 42709, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 42709
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700451595_FLI.pep
US-10-425-114-42709

Query Match          54.2%; Score 39; DB 15; Length 133;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRID 13
Db 117 KFYDSKRSD 125

RESULT 24
US-10-425-114-62912
; Sequence 62912, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62912
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3067-050-C2_FLI.pep
US-10-425-114-62912

Query Match      54.2%; Score 39; DB 15; Length 136;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      5 KFYDNKRID 13
      ||||:||||
Db      120 KFYDSKRS 128

RESULT 25
US-10-425-114-72110
; Sequence 72110, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72110
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB36-012-F10_FLI.pep
US-10-425-114-72110

Query Match      54.2%; Score 39; DB 15; Length 138;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      5 KFYDNKRID 13
      ||||:||||
Db      122 KFYDSKRS 130

RESULT 26
US-10-425-114-48114
; Sequence 48114, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48114
; LENGTH: 146
; TYPE: PRT
```

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-012-All_FLI.pep
US-10-425-114-48114

Query Match      54.2%; Score 39; DB 15; Length 146;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      5 KFYDNKRID 13
      ||||:||||
Db      130 KFYDSKRS 138

RESULT 27
US-09-893-737-50
; Sequence 50, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-50

Query Match      54.2%; Score 39; DB 9; Length 369;
Best Local Similarity 53.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      1 RSDYKFYDNKRID 13
      ||||:||||
Db      212 RVDYVFDNFR 224

RESULT 28
US-10-437-963-104111
; Sequence 104111, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 104111
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101476C.1.pep
US-10-437-963-104111

Query Match      54.2%; Score 39; DB 16; Length 636;
Best Local Similarity 77.8%; Pred. No. 5.5e+02;
```



```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_43012C.1.pep
US-10-437-963-141960

Query Match      52.8%; Score 38; DB 16; Length 187;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      5 KFYDNKRID 13
       :|||||:|
Db      176 KYDNKLLD 184

RESULT 32
US-10-282-122A-51668
; Sequence 51668, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51668
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51668

Query Match      52.8%; Score 38; DB 15; Length 207;
Best Local Similarity 46.2%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      1 RSDYKFFYDNKRID 13
       :|||||:|
Db      167 KGDYKLYSKKEIE 179

us-09-719-379a-3.rapb

RESULT 33
US-10-424-599-248762
; Sequence 248762, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 248762
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66661C.1.pep
US-10-424-599-248762

Query Match      52.8%; Score 38; DB 15; Length 212;
Best Local Similarity 77.8%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 SDYKFFYDNK 10
       :|||||:|
Db      161 SDYKFFYSK 169

RESULT 34
US-10-437-963-193680
; Sequence 193680, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 193680
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_89797C.1.pep
US-10-437-963-193680

Query Match      52.8%; Score 38; DB 16; Length 309;
Best Local Similarity 66.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 RSDYKFFYDN 9
       :|||||:|
Db      68 RGDYDFYEN 76

RESULT 35
US-10-369-493-23664
; Sequence 23664, Application US/10369493
; Publication No. US20030233675A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23664
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-23664

Query Match 52.8%; Score 38; DB 14; Length 715;
Best Local Similarity 58.3%; Pred. No. 9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
| | | | |
Db 575 REDPEFYENMRI 586

RESULT 36
US-09-853-533A-10
; Sequence 10, Application US/09853533A
; Patent No. US20020103362A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Company
; APPLICANT: Isaac, Barbara
; APPLICANT: Krieger, Elysia
; APPLICANT: Mettus, Anne-Marie
; APPLICANT: Moshiri, Farhad
; APPLICANT: Sivasubramaniam, Sakuntala
; TITLE OF INVENTION: POLYPEPTIDE COMPOSITIONS TOXIC TO ANTHONOMUS INSECTS, AND METHODS
; FILE REFERENCE: 38-21(51932)B
; CURRENT APPLICATION NUMBER: US/09/853,533A
; CURRENT FILING DATE: 2001-03-11
; PRIOR APPLICATION NUMBER: 60/204,367
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-09-853-533A-10

Query Match 52.8%; Score 38; DB 9; Length 722;
Best Local Similarity 41.7%; Pred. No. 9.1e+02;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
| | | | |
Db 611 KGDYVFYDSRHV 622

RESULT 37
US-10-360-899-10
; Sequence 10, Application US/10360899
; Publication No. US20030229919A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Company
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Krieger, Elysia K.
; APPLICANT: Mettus, Anne-Marie Light
; APPLICANT: Moshiri, Farhad
```

```
; APPLICANT: Sivasubramaniam, Sakuntala
; TITLE OF INVENTION: POLYPEPTIDE COMPOSITIONS TOXIC TO ANTHONOMUS INSECTS, AND METHODS
; FILE REFERENCE: 38-21(51932)C
; CURRENT APPLICATION NUMBER: US/10/360,899
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/204,367
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: 09/853,533
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-10-360-899-10

Query Match 52.8%; Score 38; DB 14; Length 722;
Best Local Similarity 41.7%; Pred. No. 9.1e+02;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
| | | | |
Db 611 KGDYVFYDSRHV 622

RESULT 38
US-10-782-141-19
; Sequence 19, Application US/10782141
; Publication No. US20040197917A1
; GENERAL INFORMATION:
; APPLICANT: Carozzi, Nadine
; APPLICANT: Hargiss, Tracy
; APPLICANT: Koziel, Michael G.
; APPLICANT: Duck, Nicholas B.
; APPLICANT: Carr, Brian
; TITLE OF INVENTION: AXMI-014, A Delta-Endotoxin Gene and
; TITLE OF INVENTION: Methods for Its Use
; FILE REFERENCE: 045600/274143
; CURRENT APPLICATION NUMBER: US/10/782,141
; CURRENT FILING DATE: 2004-02-20
; PRIOR APPLICATION NUMBER: 60/448,632
; PRIOR FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-10-782-141-19

Query Match 52.8%; Score 38; DB 17; Length 722;
Best Local Similarity 41.7%; Pred. No. 9.1e+02;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
| | | | |
Db 611 KGDYVFYDSRHV 622

RESULT 39
US-09-908-193-35
; Sequence 35, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
```

; CURRENT APPLICATION NUMBER: US/09/908,193  
; CURRENT FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: 60/220,273  
; PRIOR FILING DATE: 2000-07-24  
; PRIOR APPLICATION NUMBER: 60/221,650  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/221,233  
; PRIOR FILING DATE: 2000-07-27  
; PRIOR APPLICATION NUMBER: 60/220,912  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/218,875  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/218,870  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/218,901  
; PRIOR FILING DATE: 2000-07-18  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 35  
; LENGTH: 788  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-908-193-35

Query Match 52.8%; Score 38; DB 9; Length 788;  
Best Local Similarity 50.0%; Pred. No. 9.9e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SDYKFYDNKRID 13  
:|||||:|:|  
Db 399 NDYPCGNKKVD 410

RESULT 40  
US-10-369-493-3803  
; Sequence 3803, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3803  
; LENGTH: 790  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-10-369-493-3803

Query Match 52.8%; Score 38; DB 14; Length 790;  
Best Local Similarity 72.7%; Pred. No. 9.9e+02;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 DYKFYDNKRID 13  
| |||||  
Db 298 DLKFYDVVRGD 308

Search completed: November 24, 2004, 10:00:14  
Job time : 50.0326 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:14:14 : Search time 11.4457 Seconds  
(without alignments)  
75.324 Million cell updates/sec

Title: US-09-719-379A-3

Perfect score: 72

Sequence: 1 RSDYKFYDNKRID 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 43    | 59.7        | 18     | 2  | US-08-460-502-1      |
| 2          | 43    | 59.7        | 18     | 4  | US-09-148-711A-1     |
| 3          | 43    | 59.7        | 40     | 2  | US-08-460-502-10     |
| 4          | 43    | 59.7        | 40     | 4  | US-09-148-711A-10    |
| 5          | 43    | 59.7        | 359    | 1  | US-08-457-997B-2     |
| 6          | 43    | 59.7        | 359    | 3  | US-08-467-722A-2     |
| 7          | 43    | 59.7        | 359    | 4  | US-09-451-184-2      |
| 8          | 42    | 58.3        | 338    | 1  | US-08-210-394-1      |
| 9          | 41    | 56.9        | 50     | 4  | US-09-270-767-36319  |
| 10         | 41    | 56.9        | 50     | 4  | US-09-270-767-51536  |
| 11         | 40    | 55.6        | 71     | 4  | US-09-328-352-7305   |
| 12         | 40    | 55.6        | 437    | 4  | US-09-711-164-440    |
| 13         | 39    | 54.2        | 341    | 4  | US-09-248-796A-16206 |
| 14         | 39    | 54.2        | 388    | 4  | US-09-248-796A-23255 |
| 15         | 39    | 54.2        | 1131   | 4  | US-09-248-796A-19110 |
| 16         | 39    | 54.2        | 1221   | 4  | US-09-107-532A-3959  |
| 17         | 38    | 52.8        | 107    | 4  | US-09-248-796A-21288 |
| 18         | 38    | 52.8        | 623    | 4  | US-09-248-796A-14256 |
| 19         | 38    | 52.8        | 715    | 2  | US-08-849-212-6      |
| 20         | 38    | 52.8        | 722    | 1  | US-08-158-232-51     |
| 21         | 38    | 52.8        | 722    | 2  | US-08-611-928-51     |
| 22         | 38    | 52.8        | 722    | 3  | US-09-173-891-51     |
| 23         | 38    | 52.8        | 722    | 4  | US-09-853-533A-10    |
| 24         | 38    | 52.8        | 740    | 4  | US-09-489-039A-13001 |
| 25         | 38    | 52.8        | 843    | 3  | US-09-235-451-25     |
| 26         | 38    | 52.8        | 843    | 4  | US-09-978-303-25     |
| 27         | 37.5  | 52.1        | 1091   | 3  | US-08-633-768A-2     |

|    |      |      |      |   |                      |                   |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 37.5 | 52.1 | 1091 | 4 | US-09-280-197-2      | Sequence 2, Appli |
| 29 | 37   | 51.4 | 263  | 2 | US-08-809-267-4      | Sequence 4, Appli |
| 30 | 37   | 51.4 | 263  | 5 | PCT-US95-13662A-4    | Sequence 4, Appli |
| 31 | 37   | 51.4 | 675  | 3 | US-09-134-001C-4547  | Sequence 4547, Ap |
| 32 | 36   | 50.0 | 71   | 4 | US-09-134-001C-6315  | Sequence 6315, Ap |
| 33 | 36   | 50.0 | 117  | 4 | US-09-328-352-4998   | Sequence 4998, Ap |
| 34 | 36   | 50.0 | 123  | 1 | US-07-893-929A-10    | Sequence 10, Appl |
| 35 | 36   | 50.0 | 123  | 5 | PCT-US92-10344-10    | Sequence 10, Appl |
| 36 | 36   | 50.0 | 153  | 1 | US-07-695-564-9      | Sequence 9, Appli |
| 37 | 36   | 50.0 | 153  | 1 | US-08-241-387-9      | Sequence 9, Appli |
| 38 | 36   | 50.0 | 246  | 4 | US-09-107-532A-5102  | Sequence 5102, Ap |
| 39 | 36   | 50.0 | 298  | 2 | US-08-838-543-5      | Sequence 5, Appli |
| 40 | 36   | 50.0 | 320  | 2 | US-08-245-511-4      | Sequence 4, Appli |
| 41 | 36   | 50.0 | 320  | 2 | US-08-600-993A-4     | Sequence 4, Appli |
| 42 | 36   | 50.0 | 385  | 4 | US-09-311-021-78     | Sequence 78, Appl |
| 43 | 36   | 50.0 | 451  | 4 | US-09-107-532A-5352  | Sequence 5352, Ap |
| 44 | 36   | 50.0 | 469  | 4 | US-09-248-796A-16207 | Sequence 16207, A |
| 45 | 36   | 50.0 | 505  | 4 | US-09-593-110-4228   | Sequence 4228, Ap |

## ALIGNMENTS

RESULT 1  
US-08-460-502-1  
; Sequence 1, Application US/08460502  
; Patent No. 5843464  
; GENERAL INFORMATION:  
; APPLICANT: Bakaletz, Lauren O.  
; APPLICANT: Kaumaya, Parvin T.  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calfee, Halter and Griswold  
; STREET: 800 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2688  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,502  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goltick, Mary E.  
; REGISTRATION NUMBER: 34,829  
; REFERENCE/DOCKET NUMBER: 22727/00120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 622-8458  
; TELEFAX: (216) 241-0816  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-460-502-1

Query Match 59.7%; Score 43; DB 2; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.57;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9  
Db 1 RSDYKFYED 9

```
RESULT 2
US-09-148-711A-1
; Sequence 1, Application US/09148711A
; Patent No. 6436405
; GENERAL INFORMATION:
; APPLICANT: The Ohio State University
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides
; FILE REFERENCE: 18525-04010
; CURRENT APPLICATION NUMBER: US/09/148,711A
; CURRENT FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 08/460,502
; PRIOR FILING DATE: 1995-06-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Haemophilus influenza
US-09-148-711A-1
Query Match          59.7%; Score 43; DB 4; Length 18;
Best Local Similarity 77.8%; Pred. No. 0.57;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9
| | | | | | | |
Db 1 RSDYKFYED 9

RESULT 3
US-08-460-502-10
; Sequence 10, Application US/08460502
; Patent No. 5843464
; GENERAL INFORMATION:
; APPLICANT: Bakaletz, Lauren O.
; APPLICANT: Kaumaya, Parvin T.
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,502
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 22727/00120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8458
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-460-502-10
Query Match          59.7%; Score 43; DB 2; Length 40;
Best Local Similarity 77.8%; Pred. No. 1.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 RSDYKFYDN 9
| | | | | | | |
Db 1 RSDYKFYED 9

RESULT 4
US-09-148-711A-10
; Sequence 10, Application US/09148711A
; Patent No. 6436405
; GENERAL INFORMATION:
; APPLICANT: The Ohio State University
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides
; FILE REFERENCE: 18525-04010
; CURRENT APPLICATION NUMBER: US/09/148,711A
; CURRENT FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 08/460,502
; PRIOR FILING DATE: 1995-06-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 40
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-148-711A-10
Query Match          59.7%; Score 43; DB 4; Length 40;
Best Local Similarity 77.8%; Pred. No. 1.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9
| | | | | | | |
Db 1 RSDYKFYED 9

RESULT 5
US-08-457-997B-2
; Sequence 2, Application US/08457997B
; Patent No. 5766608
; GENERAL INFORMATION:
; APPLICANT: Kolattukudy, P. E.
; TITLE OF INVENTION: Otitis Media Vaccine
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter and Griswold
; STREET: Suite 1800 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,997B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 22727/00102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8458
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-457-997B-2
```



Query Match 59.7%; Score 43; DB 1; Length 359;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9  
Db 138 RSDYKFYED 146

## RESULT 6

US-08-467-722A-2

; Sequence 2, Application US/08467722A  
; Patent No. 6030626  
; GENERAL INFORMATION:  
; APPLICANT: Kolattukudy, P. E.  
; TITLE OF INVENTION: Otitis Media Vaccine  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calfee, Halter and Griswold  
; STREET: Suite 1800 800 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2688

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/467,722A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Golrick, Mary E.  
; REGISTRATION NUMBER: 34,829  
; REFERENCE/DOCKET NUMBER: 22727/00102  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 622-8458  
; TELEFAX: (216) 241-0816  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 359 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-467-722A-2

Query Match 59.7%; Score 43; DB 3; Length 359;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9  
Db 138 RSDYKFYED 146

## RESULT 7

US-09-451-184-2

; Sequence 2, Application US/094511184  
; Patent No. 6562349  
; GENERAL INFORMATION:  
; APPLICANT: Kolattukudy, P. E.  
; TITLE OF INVENTION: Otitis Media Vaccine  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calfee, Halter and Griswold  
; STREET: 1400 McDonald Investment Center,  
; STREET: 800 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.

; ZIP: 44114-2688  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09/451,184  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Docherty, Pamela A.  
; REGISTRATION NUMBER: 40,591  
; REFERENCE/DOCKET NUMBER: 24547/04000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 622-8416  
; TELEFAX: (216) 241-0816  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 359 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-451-184-2

Query Match 59.7%; Score 43; DB 4; Length 359;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9  
Db 138 RSDYKFYED 146

## RESULT 8

US-08-210-394-1  
; Sequence 1, Application US/08210394  
; Patent No. 5770213  
; GENERAL INFORMATION:  
; APPLICANT: Zlotnick Dr., Gary W.  
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus  
; TITLE OF INVENTION: Influenzae P5 Protein as a Vaccine  
; TITLE OF INVENTION: Haemophilus Influenzae Strain  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: US  
; ZIP: 07470-8426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/210,394  
; FILING DATE: 07-MAR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J  
; REFERENCE/DOCKET NUMBER: 32,144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201/831-3246  
; TELEFAX: 201/831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: protein  
; MOLECULE TYPE: protein

## US-08-210-394-1

Query Match 58.3%; Score 42; DB 1; Length 338;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYD 8  
| | | | |  
Db 119 RSDYKFYE 126

## RESULT 9

US-09-270-767-36319  
; Sequence 36319, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 36319  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-36319

Query Match 56.9%; Score 41; DB 4; Length 50;  
Best Local Similarity 61.5%; Pred. No. 3.6;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13  
| | | | |  
Db 5 RFDHKHYDNRND 17

## RESULT 10

US-09-270-767-51536  
; Sequence 51536, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51536  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-51536

Query Match 56.9%; Score 41; DB 4; Length 50;  
Best Local Similarity 61.5%; Pred. No. 3.6;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13  
| | | | |  
Db 5 RFDHKHYDNRND 17

## RESULT 11

US-09-328-352-7305  
; Sequence 7305, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Brston et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7305  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7305

Query Match 55.6%; Score 40; DB 4; Length 71;  
Best Local Similarity 56.3%; Pred. No. 7.8;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRID 13  
| | | | |  
Db 16 SSYKFYFNKKIE 27

## RESULT 12

US-09-711-164-440  
; Sequence 440, Application US/09711164  
; Patent No. 6589738  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
; FILE REFERENCE: ELITRA.008A  
; CURRENT APPLICATION NUMBER: US/09/711,164  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-9  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 440  
; LENGTH: 437  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-711-164-440

Query Match 55.6%; Score 40; DB 4; Length 437;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
| | | | |  
Db 367 DYKYDN 373

## RESULT 13

US-09-248-796A-16206  
; Sequence 16206, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 16206  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-16206

Query Match 54.2%; Score 39; DB 4; Length 341;

```
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 FYDNKRID 13
Db 320 FYDNKRVD 327

RESULT 14
US-09-248-796A-23255
; Sequence 23255, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 23255
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-23255

Query Match 54.2%; Score 39; DB 4; Length 388;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 12
Db 302 DYVLDNKKL 311

RESULT 15
US-09-248-796A-19110
; Sequence 19110, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19110
; LENGTH: 1131
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19110

Query Match 54.2%; Score 39; DB 4; Length 1131;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 12
Db 169 DKYIENKRL 178

RESULT 16
US-09-107-532A-3959
; Sequence 3959, Application US/09107532A
```

```
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3959:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1221
; SEQUENCE DESCRIPTION: SEQ ID NO: 3959:
US-09-107-532A-3959

Query Match 54.2%; Score 39; DB 4; Length 1221;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 FYDNKRID 13
Db 693 FYDNKRVD 700

RESULT 17
US-09-248-796A-21288
; Sequence 21288, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
```

```
; SEQ ID NO 21288
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21288

Query Match      52.8%; Score 38; DB 4; Length 107;
Best Local Similarity 55.6%; Pred. No. 26;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KYFDNKRRI 12
   :|||||:
Db 53 FKFDNQKL 61

RESULT 18
US-09-248-796A-14256
; Sequence 14256, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14256
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14256

Query Match      52.8%; Score 38; DB 4; Length 623;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SDYKPYD 8
   :||||:
Db 56 SDYKYD 62

RESULT 19
US-08-849-212-6
; Sequence 6, Application US/08849212
; Patent No. 5827698
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, YOSHIMI
; APPLICANT: SUZUKI, TOMOKO
; APPLICANT: KOJIMA, HIROYUKI
; TITLE OF INVENTION: NOVEL LYSINE DECARBOXYLASE GENE AND
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,212
; FILING DATE: 09-JUN-1997
```

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6/306386
; FILING DATE: 09-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-856-0PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-849-212-6

Query Match      52.8%; Score 38; DB 2; Length 715;
Best Local Similarity 58.3%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKRRI 12
   :||:|:|:|
Db 575 REDPEFYENMRI 586

RESULT 20
US-08-158-232-51
; Sequence 51, Application US/08158232
; Patent No. 5596071
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Kennedy, M. Keith
; APPLICANT: Randall, John Brooks
; APPLICANT: Meier, Henry
; APPLICANT: Uick, Heidi Jane
; APPLICANT: Foncetrada, Luis
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Schwab, George E.
; APPLICANT: Fu, Jenny
; TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
; TITLE OF INVENTION: Against Hymenopteran Pests
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,232
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/887,980
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/797,645
; FILING DATE: 25-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/703,977
; FILING DATE: 22-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
```

REFERENCE/DOCKET NUMBER: M/SCU104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 722 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-158-232-51

Query Match 52.8%; Score 38; DB 1; Length 722;  
Best Local Similarity 41.7%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDYDNKRI 12  
Db 611 KGDYVFYDSRHV 622

## RESULT 21

US-08-611-928-51  
Sequence 51, Application US/08611928  
Patent No. 5824792  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Fonceerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCU104.C1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 722 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-928-51

Query Match 52.8%; Score 38; DB 2; Length 722;  
Best Local Similarity 41.7%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDYDNKRI 12  
Db 611 KGDYVFYDSRHV 622

## RESULT 22

US-09-173-891-51  
Sequence 51, Application US/09173891  
Patent No. 6077937  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Fonceerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,891  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/158,232  
FILING DATE:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCU104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:  
LENGTH: 722 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-173-891-51

Query Match 52.8%; Score 38; DB 3; Length 722;  
Best Local Similarity 41.7%; Pred. No. 2e+02;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKR1 12  
: || ||| : :  
Db 611 KGDYVFDNRHV 622

## RESULT 23

US-09-853-533A-10  
Sequence 10, Application US/09853533A  
Patent No. 6541448

## GENERAL INFORMATION:

APPLICANT: Monsanto Company  
APPLICANT: Isaac, Barbara  
APPLICANT: Krieger, Elysia  
APPLICANT: Mettus, Anne-Marie  
APPLICANT: Moshiri, Farhad  
APPLICANT: Sivasubramanian, Sakuntala  
TITLE OF INVENTION: POLYPEPTIDE COMPOSITIONS TOXIC TO ANTHONOMUS INSECTS, AND METHODS  
FILE REFERENCE: 38-21(51932)B  
CURRENT APPLICATION NUMBER: US/09/853.533A  
CURRENT FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: 60/204,367  
PRIOR FILING DATE: 2000-05-15  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 10  
LENGTH: 722  
TYPE: PRT  
ORGANISM: Bacillus thuringiensis  
US-09-853-533A-10

Query Match 52.8%; Score 38; DB 4; Length 722;  
Best Local Similarity 41.7%; Pred. No. 2e+02;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKR1 12  
: || ||| : :  
Db 611 KGDYVFDNRHV 622

## RESULT 24

US-09-489-039A-13001  
Sequence 13001, Application US/09489039A  
Patent No. 6610836

## GENERAL INFORMATION:

APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489.039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 13001  
LENGTH: 740  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13001

Query Match 52.8%; Score 38; DB 4; Length 740;  
Best Local Similarity 58.3%; Pred. No. 2e+02;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 RSDYKFDNKR1 12  
: || ||| : :  
Db 600 REDPEFYENMRI 611

## RESULT 25

US-09-235-451-25  
Sequence 25, Application US/09235451  
GENERAL INFORMATION:

APPLICANT: Julius, David J.  
APPLICANT: Caterina, Michael J.  
APPLICANT: Brake, Anthony J.  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING  
TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED  
FILE REFERENCE: 9076/084CIP  
CURRENT APPLICATION NUMBER: US/09/235,451  
CURRENT FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/072,151  
PRIOR FILING DATE: 1998-01-22  
PRIOR APPLICATION NUMBER: 08/915,461  
PRIOR FILING DATE: 1997-08-20  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 25  
LENGTH: 843  
TYPE: PRT  
ORGANISM: chicken  
US-09-235-451-25

Query Match 52.8%; Score 38; DB 3; Length 843;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 YKFDNKR1 12  
: ||| : ||  
Db 116 FKPYDRRR1 124

## RESULT 26

US-09-978-303-25  
Sequence 25, Application US/09978303  
Patent No. 6790629

## GENERAL INFORMATION:

APPLICANT: Julius, David J.  
APPLICANT: Caterina, Michael J.  
APPLICANT: Brake, Anthony J.  
TITLE OF INVENTION: Nucleic acid sequences encoding  
TITLE OF INVENTION: capsaicin receptor and capsaicin receptor-related  
FILE REFERENCE: UCAL084CON  
CURRENT APPLICATION NUMBER: US/09/978,303  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/235,451  
PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/072,151  
PRIOR FILING DATE: 1998-01-22  
PRIOR APPLICATION NUMBER: 08/915,461  
PRIOR FILING DATE: 1997-08-20  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 25  
LENGTH: 843  
TYPE: PRT  
ORGANISM: chicken  
US-09-978-303-25

Query Match 52.8%; Score 38; DB 4; Length 843;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

Qy      4 YKFDNKR1 12
Db      116 FKFDYRR1 124

RESULT 27
US-08-633-768A-2
; Sequence 2, Application US/08633768A
; Patent No. 6013504
; GENERAL INFORMATION:
; APPLICANT: YU, SHUKUN
; APPLICANT: BOJSEN, KIRSTEN
; APPLICANT: KRAGH, KARSTEN
; APPLICANT: BOJKO, MAJA
; APPLICANT: NIELSEN, JOHN
; APPLICANT: MARCUSSON, JAN
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633.768A
; FILING DATE: 02-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9321301.5
; FILING DATE: 15-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0U7.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1091 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-633-768A-2

Query Match      52.1%; Score 37.5; DB 3; Length 1091;
Best Local Similarity 61.5%; Pred. No. 3.7e+02;
Matches      8; Conservative      1; Mismatches      1; Indels      3; Gaps      1;

Qy      4 YKF---YDNKRID 13
Db      894 YKFGPDYDKRLD 906

RESULT 28
US-08-280-197-2
; Sequence 2, Application US/09280197
; Patent No. 6632643
; GENERAL INFORMATION:
; APPLICANT: Yu, Shukun
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Kragh, Karsten
; APPLICANT: Bojko, Maja

```

```

; APPLICANT: Nielsen, John
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: USE OF "-1,4-GLUCAN LYASE FOR PREPARATION OF
; TITLE OF INVENTION: 1,5-D-ANHYDROFRUCTOSE
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/280,197
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/633,719
; FILING DATE: July 8, 1996
; APPLICATION NUMBER: PCT/EP94/03397
; FILING DATE: OCT-15-1994
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0U5.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1091 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-280-197-2

Query Match      52.1%; Score 37.5; DB 4; Length 1091;
Best Local Similarity 61.5%; Pred. No. 3.7e+02;
Matches      8; Conservative      1; Mismatches      1; Indels      3; Gaps      1;

Qy      4 YKF---YDNKRID 13
Db      894 YKFGPDYDKRLD 906

RESULT 29
US-08-809-267-4
; Sequence 4, Application US/08809267
; Patent No. 5861296
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Barton E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,267
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 433
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13662
; FILING DATE:
; APPLICATION NUMBER: US 08/329,721
; FILING DATE: 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-809-267-4

Query Match 51.4%; Score 37; DB 2; Length 263;
Best Local Similarity 54.5%; Pred. No. 1e+02; Indels 2; Mismatches 3; Gaps 0;
Matches 6; Conservative 3;

Qy 3 DYKFDNKRID 13
||| :||:|
Db 177 DYKKNKKVD 187

RESULT 30
PCT-US95-13662A-4
; Sequence 4, Application PC/TUS9513662A
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Barton E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13662A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,721
; FILING DATE: 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
```

```
;
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US95-13662A-4

Query Match 51.4%; Score 37; DB 5; Length 263;
Best Local Similarity 54.5%; Pred. No. 1e+02; Indels 2; Mismatches 3; Gaps 0;
Matches 6; Conservative 3;

Qy 3 DYKFDNKRID 13
||| :||:|
Db 177 DYKKNKKVD 187

RESULT 31
US-09-134-001C-4547
; Sequence 4547, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4547
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
; US-09-134-001C-4547

Query Match 51.4%; Score 37; DB 3; Length 675;
Best Local Similarity 50.0%; Pred. No. 2.7e+02; Indels 4; Gaps 1;
Matches 8; Conservative 2; Mismatches 2;

Qy 2 SDYKFYD---NKRID 13
||| ||: ||:|
Db 501 SDYFFYNAQISNKKLD 516

RESULT 32
US-09-134-000C-6315
; Sequence 6315, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6315
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
```



US-09-134-000C-6315

Query Match 50.0%; Score 36; DB 4; Length 71;  
Best Local Similarity 66.7%; Pred. No. 37;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 YKFDNKR 12  
| | | | |  
Db 45 YSFYKRNKI 53

RESULT 33

US-09-328-352-4998  
; Sequence 4998, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4998  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4998

Query Match 50.0%; Score 36; DB 4; Length 117;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKR 11  
: : : : :  
Db 8 KGEYKFNNK 18

RESULT 34

US-07-893-929A-10  
; Sequence 10, Application US/07893929A  
; Patent No. 5336667  
; GENERAL INFORMATION:  
; APPLICANT: Kirby, Edward P.  
; APPLICANT: Peng, Man-ling  
; TITLE OF INVENTION: Alboaggregins: Platelet  
; TITLE OF INVENTION: Agonists Which Bind To Platelet  
; TITLE OF INVENTION: Membrane Glycoprotein Ib  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Temple University - Of The Common-  
; ADDRESSEE: wealth System of Higher Education  
; STREET: 406 University Services Building  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: U.S.A.  
; ZIP: 19122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/893,929A  
; FILING DATE: 19920605  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/803,630  
; FILING DATE: December 3, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monaco, Daniel A.  
; REGISTRATION NUMBER: 30,480  
; REFERENCE/DOCKET NUMBER: 6056-126 (CIP) 1

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-8383  
; TELEFAX: (215) 568-5549  
; TELEX: No. 5336667e  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 123 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
US-07-893-929A-10

Query Match 50.0%; Score 36; DB 1; Length 123;  
Best Local Similarity 66.7%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 11  
| | | : | : |  
Db 89 DYKAWDNER 97

RESULT 35

PCT-US92-10344-10  
; Sequence 10, Application PC/TUS9210344  
; GENERAL INFORMATION:  
; APPLICANT: Kirby, Edward P.  
; APPLICANT: Peng, Man-ling  
; TITLE OF INVENTION: Alboaggregins: Platelet  
; TITLE OF INVENTION: Agonists Which Bind To Platelet  
; TITLE OF INVENTION: Membrane Glycoprotein Ib  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Temple University - Of The Common-  
; ADDRESSEE: wealth System of Higher Education  
; STREET: 406 University Services Building  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: U.S.A.  
; ZIP: 19122

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/10344  
; FILING DATE: 19921201  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/803,630  
; FILING DATE: December 3, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monaco, Daniel A.  
; REGISTRATION NUMBER: 30,480  
; REFERENCE/DOCKET NUMBER: 6056-126 (CIP) 1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-8383  
; TELEFAX: (215) 568-5549  
; TELEX: None  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 123 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
PCT-US92-10344-10

Query Match 50.0%; Score 36; DB 5; Length 123;  
Best Local Similarity 66.7%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 11  
| | | : | : |  
Db 89 DYKAWDNER 97

RESULT 36  
US-07-695-564-9  
; Sequence 9, Application US/07695564  
; Patent No. 5310874  
; GENERAL INFORMATION:  
; APPLICANT: Tamura, Richard N.  
; APPLICANT: Quaranta, Vito  
; TITLE OF INVENTION: INTEGRIN ALPHA SUBUNIT CYTOPLASMIC  
; TITLE OF INVENTION: DOMAIN POLYPEPTIDES, ANTIBODIES AND METHODS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thomas Fitting  
; STREET: 11300 Sorrento Valley Road, Suite 200  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/695,564  
; FILING DATE: 19910503  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: SCR0377P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-1555  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 153 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 1..153  
; OTHER INFORMATION: /note= "SEQ ID NO:9 is the 153  
; OTHER INFORMATION: amino acid sequence predicted from the product  
; OTHER INFORMATION: which results from amplification of the mouse  
; OTHER INFORMATION: ALPHA 3B cDNA with primers 2032/2033."  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 108..112  
; OTHER INFORMATION: /note= "The cytoplasmic sequence  
; OTHER INFORMATION: CDFFK begins at amino acid position 108."  
US-07-695-564-9

Query Match 50.0%; Score 36; DB 1; Length 153;  
Best Local Similarity 54.5%; Pred. No. 83;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKFYDNKRID 13  
||| :| :| :|  
Db 33 DYKDFDRVRD 43

RESULT 37  
US-08-241-387-9  
; Sequence 9, Application US/08241387  
; Patent No. 5589570  
; GENERAL INFORMATION:

; APPLICANT: Tamura, Richard N.  
; APPLICANT: Quaranta, Vito  
; TITLE OF INVENTION: INTEGRIN ALPHA SUBUNIT CYTOPLASMIC  
; TITLE OF INVENTION: DOMAIN POLYPEPTIDES, ANTIBODIES AND METHODS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute  
; STREET: 10666 No. 5589570th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/241,387  
; FILING DATE: 10-MAY-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 07/695,564  
; FILING DATE: 03-MAY-1004  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSRI241.0D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-554-2937  
; TELEFAX: 619-554-6312  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 153 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 1..153  
; OTHER INFORMATION: /note= "SEQ ID NO:9 is the 153  
; OTHER INFORMATION: amino acid sequence predicted from the product  
; OTHER INFORMATION: which results from amplification of the mouse  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 108..112  
; OTHER INFORMATION: /note= "The cytoplasmic sequence  
; OTHER INFORMATION: CDFFK begins at amino acid position 108."  
US-08-241-387-9

Query Match 50.0%; Score 36; DB 1; Length 153;  
Best Local Similarity 54.5%; Pred. No. 83;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKFYDNKRID 13  
||| :| :| :|  
Db 33 DYKDFDRVRD 43

RESULT 38  
US-09-107-532A-5102  
; Sequence 5102, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 5102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...246
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 5102:
;
; US-09-107-532A-5102
;
; Query Match 50.0%; Score 36; DB 4; Length 246;
; Best Local Similarity 62.5%; Pred. No. 1.4e+02;
; Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 RSDYKFYD 8
; Db 224 KADYEFYD 231
;
; RESULT 39
; US-08-838-543-5
; Sequence 5, Application US/08838543
; Patent No. 5994623
;
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, ENNO
; APPLICANT: BROGLIE, KAREN E.
; TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 INCH DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,543
; FILING DATE:
```

```
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: P-41,173
; REFERENCE/DOCKET NUMBER: BB-1101
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0184
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 298 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-08-838-543-5
;
; Query Match 50.0%; Score 36; DB 2; Length 298;
; Best Local Similarity 53.8%; Pred. No. 1.7e+02;
; Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
;
; Qy 1 RSDYKFYDNRKID 13
; Db 76 RESPKYDIVRID 88
;
; RESULT 40
; US-08-245-511-4
; Sequence 4, Application US/08245511
; Patent No. 5928900
;
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 424
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-245-511-4
;
; Query Match 50.0%; Score 36; DB 2; Length 320;
```

Best Local Similarity 45.5%; Pred. NO. 1.8e+02;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 3 DYKFYDNKRID 13  
|::|::|  
Db 43 DHRFFDHRGID 53

Search completed: November 24, 2004, 09:31:40  
Job time : 12.4457 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:04 ; Search time 59.7717 Seconds  
(without alignments)  
125.141 Million cell updates/sec

Title: US-09-719-379A-3  
Perfect score: 72  
Sequence: 1 RSDYKFYDNKRID 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | ID         | Description         |
|------------|-------|---------------|--------|------------|---------------------|
| 1          | 51    | 70.8          | 1047   | Q7RKV1     | Q7rkvl plasmodium   |
| 2          | 46    | 63.9          | 360    | O86254     | O86254 haemophilus  |
| 3          | 44    | 61.1          | 212    | Q6CWX3     | Q6cwx3 kiuyveromyc  |
| 4          | 43    | 59.7          | 76     | Q6Z5F0     | Q6z5f0 oryza sativ  |
| 5          | 43    | 59.7          | 76     | BAD10106   | BAD10106 oryza sat  |
| 6          | 43    | 59.7          | 351    | Q6GUB7     | Q6gub7 pasteurella  |
| 7          | 43    | 59.7          | 353    | OM52_HAEIN | P38368 haemophilus  |
| 8          | 43    | 59.7          | 359    | OM53_HAEIN | P45996 haemophilus  |
| 9          | 43    | 59.7          | 404    | Q6XE78     | Q6xe78 uncultured   |
| 10         | 43    | 59.7          | 404    | AAP70366   | AAP70366 gamma-pro  |
| 11         | 43    | 59.7          | 451    | Q6XE55     | Q6xe55 uncultured   |
| 12         | 43    | 59.7          | 451    | AAP49314   | AAP49314 gamma-pro  |
| 13         | 43    | 59.7          | 451    | AAP70389   | AAP70389 gamma-pro  |
| 14         | 43    | 59.7          | 1905   | Y659_PAGMU | Q9cmz1 pasteurella  |
| 15         | 43    | 59.7          | 2616   | Q8IIIG1    | Q8lig1 plasmodium   |
| 16         | 42    | 58.3          | 124    | Q6MS53     | Q6ms53 mycoplasma   |
| 17         | 42    | 58.3          | 124    | Cae77537   | Cae77537 mycoplasma |
| 18         | 42    | 58.3          | 248    | Q8IK12     | Q8ik12 plasmodium   |
| 19         | 42    | 58.3          | 286    | Q9UWV0     | Q9uwv0 sulfolobus   |
| 20         | 42    | 58.3          | 360    | Q7RNX0     | Q7rnx0 plasmodium   |
| 21         | 42    | 58.3          | 517    | Q8A4H6     | Q8a4h6 bacteroides  |
| 22         | 42    | 58.3          | 870    | Q7VXJ8     | Q7vxj8 bordetella   |
| 23         | 42    | 58.3          | 870    | Q7W8Y6     | Q7w8y6 bordetella   |
| 24         | 42    | 58.3          | 870    | Q7WKC9     | Q7wkc9 bordetella   |
| 25         | 42    | 58.3          | 903    | Q9A915     | Q9a915 caulobacter  |
| 26         | 42    | 58.3          | 1021   | Q8A3D0     | Q8a3d0 bacteroides  |
| 27         | 42    | 58.3          | 1102   | Q7YYQ2     | Q7yyq2 cryptospori  |
| 28         | 41    | 56.9          | 309    | Q6MTB3     | Q6mtb3 mycoplasma   |
| 29         | 41    | 56.9          | 359    | Cae77123   | Cae77123 mycoplasma |
| 30         | 41    | 56.9          | 354    | Q7RQ14     | Q7rq14 plasmodium   |
| 31         | 41    | 56.9          | 1063   | HGP1_HAEIN | P44795 haemophilus  |

|    |    |      |      |   |            |                     |
|----|----|------|------|---|------------|---------------------|
| 32 | 41 | 56.9 | 1066 | 1 | HGPC_HAEIN | Q9x442 haemophilus  |
| 33 | 41 | 56.9 | 1067 | 1 | HGPB_HAEIN | Q9xivi haemophilus  |
| 34 | 41 | 56.9 | 1084 | 1 | HGP3_HAEIN | P44836 haemophilus  |
| 35 | 40 | 55.6 | 172  | 2 | Q89T91     | Q89t91 bradyrhizob  |
| 36 | 40 | 55.6 | 318  | 2 | Q6ZEW6     | Q6zew6 synecocyst   |
| 37 | 40 | 55.6 | 318  | 2 | BAD01784   | BAD01784 synecocyst |
| 38 | 40 | 55.6 | 349  | 2 | O6GUB8     | O6gub8 pasteurella  |
| 39 | 40 | 55.6 | 353  | 2 | Q6GUB4     | Q6gub4 pasteurella  |
| 40 | 40 | 55.6 | 353  | 2 | Q6GUB5     | Q6gub5 pasteurella  |
| 41 | 40 | 55.6 | 353  | 2 | Q9CMN1     | Q9cmn1 pasteurella  |
| 42 | 40 | 55.6 | 395  | 2 | Q8A2K4     | Q8a2k4 bacteroides  |
| 43 | 40 | 55.6 | 404  | 2 | Q6MUA5     | Q6mua5 mycoplasma   |
| 44 | 40 | 55.6 | 404  | 2 | Cae76779   | Cae76779 mycoplasma |
| 45 | 40 | 55.6 | 408  | 1 | AATC_CABEL | Q22067 caenorhabdi  |

#### ALIGNMENTS

RESULT 1  
Q7RKV1  
ID Q7RKV1 PRELIMINARY; PRT; 1047 AA.  
AC Q7RKV1;  
DT 01-MAR-2004 (TREMBlrel. 26, Created)  
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Unnamed protein product.  
GN Name=PY02799;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=73239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17XNL;  
RX PubMed=12368865;  
RA Carlton J.M., Angioli S.V., Suh B.B., Koolij T.W., Partea M., Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V., Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M., Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;  
RT "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";  
RL Nature 419:512-519(2002)  
CC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.  
CC EMBL; AABL0100780; EAA22293.1; -.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 3.  
DR PROSITE; PS50005; TPR; 1.  
DR PROSITE; PS50293; TPR REGION; 1.  
SQ SEQUENCE 1047 AA; 123758 MW; 2FE31D626FD42077 CRC64;

Query Match 70.8%; Score 51; DB 2; Length 1047;

Best Local Similarity 69.2%; Pred. No. 5.9;

Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13

||||| :  
484 RSDYKIIDNKSVD 496

RESULT 2

O86254 PRELIMINARY; PRT; 360 AA.

ID O86254

AC O86254;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Outer membrane protein (Fragment).  
GN Name=omp;  
OS Haemophilus sp.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_taxID=740;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16N;  
RX MEDLINE=93081716; PubMed=9864189;  
RA Goussset N., Rosenau A., Sizaret P.Y., Quentin R.;  
RT "Nucleotide sequences of genes coding for fimbrial proteins in a  
RT cryptic genospecies of Haemophilus spp. isolated from neonatal and  
RT genital tract infections.";  
RL Infect. Immun. 67:8-15(1999).  
CC -!- SIMILARITY: Belongs to the ompA family.  
DR EMBL; AJ007317; CAA07454.1; -.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0009279; C:outer membrane (sensu Gram-negative Bacteria); IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR006664; Bac.OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR000498; OmpA\_Like.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA membrane; 1.  
DR PRINTS; PR01021; OMPADOMAIN.  
DR PRINTS; PR01022; OUTRMBRANE.  
DR ProDom; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OMPA; 1.  
FT NON\_TER 360 360  
SQ SEQUENCE 360 AA; 38415 MW; A3209155051CDD69 CRC64;  
  
Query Match 63.9%; Score 46; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 RSDYKFYD 8  
Db 140 RSDYKFYD 147  
  
RESULT 3  
Q6CWZ3 PRELIMINARY; PRT; 212 AA.  
ID Q6CWZ3  
AC Q6CWZ3 (TrEMBLrel. 28, Created)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DE Similarity.  
GN ORFNames=KLLA0B00363g;  
OS Kluyveromyces fragilis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL Y-1140;  
RG GENOLEVURES;  
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Tallia E.,  
RA Goffard N., Franchin L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Boursac S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
RA Boissarie A., Boyer J., Cattolico L., Confalonieri F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jaumiaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul L., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.P., Straub M.L., Suleau A.,  
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,

RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.L.;  
RT "Genome evolution in yeasts";  
RN Nature 430:35-44(2004).  
EN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL Y-1140;  
RA Genoscope;  
RL EMBL; CR382122; CAH01939.1; -.  
SQ SEQUENCE 212 AA; 24162 MW; B3888750EF4D6C10 CRC64;  
  
Query Match 61.1%; Score 44; DB 2; Length 212;  
Best Local Similarity 53.8%; Pred. No. 18;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 RSDYKFYDNRKID 13  
Db 19 RSTFLFYDNKQLE 31  
  
RESULT 4  
Q6Z5F0 PRELIMINARY; PRT; 76 AA.  
ID Q6Z5F0  
AC Q6Z5F0 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE Hypothetical protein B1142B04.6.  
GN Name=B1142B04.6;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005148; BADI0106.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 76 AA; 9017 MW; 91B11512447964F8 CRC64;  
  
Query Match 59.7%; Score 43; DB 2; Length 76;  
Best Local Similarity 58.3%; Pred. No. 8.9;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 RSDYKFYDNRKRI 12  
Db 47 RSSFKFYDRKAV 58  
  
RESULT 5  
BADI0106 PRELIMINARY; PRT; 76 AA.  
ID BADI0106  
AC BADI0106 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DE Hypothetical protein B1142B04.6.  
GN B1142B04.6.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC  
RT clone:B1142B04.1";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AP005148; BAD10106.1; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 9017 MW; 91B11512447964F8 CRC64;

Query Match      59.7%; Score 43; DB 2; Length 76;
Best Local Similarity 58.3%; Pred. No. 8.9;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKRID 12
Db 47 RSSFKFYDRKAV 58

RESULT 6
Q6GUB7 PRELIMINARY; PRT; 351 AA.
AC Q6GUB7;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE OmpA.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95120769;
RA Akridge H., Confer A.W., Dabo S.M.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBSJ databases.
CC -1- SIMILARITY: Belongs to the ompA family.
DR EMBL; AY643795; AAT57677.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR002368; OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA LIKE.
DR InterPro; IPR000498; OmpA_tmam.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PF01389; OmpA_membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR PRODOM; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OmpA; 1.
DR Direct protein sequencing; Outer membrane; Porin; Signal;
KW Transmembrane.
FT CHAIN 1 21 Outer membrane protein P5.
FT DISULFID 326 338 By similarity.
FT DOMAIN 272 316 OmpA-like
SQ SEQUENCE 353 AA; 37594 MW; E58A659E786D0DF7 CRC64;

Query Match      59.7%; Score 43; DB 1; Length 353;
Best Local Similarity 77.8%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDN 9
Db 132 RSDYKFDYED 140

RESULT 8
OM53 HAEIN
ID OM53 HAEIN STANDARD; PRT; 359 AA.
AC P45996;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Outer membrane protein P5 precursor (OMP P5) (Fimbrin).
GN Name=ompA; Synonym=ompP5;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NTHI 1128;
RX MEDLINE=94222575; PubMed=7909539;
RA Demaria T., Kolattukudy P.E., Murwin D., Billy J., Leake E., Lim D.,
RA Sirakova T., Bakaletz L.;
RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
RT pathogenesis of and protection against otitis media and relatedness of
RT the fimbrin subunit to outer membrane protein A.";
RL Infect. Immun. 62:2002-2020(1994).
CC -1- FUNCTION: Acts as a fimbriae subunit.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: Belongs to the ompA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; L08448; AA24959.1; -.  
DR HSRF; P02934; IEXW.  
DR InterPro; IPR006664; Bac\_OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR006690; OmpA LIKE.  
DR InterPro; IPR000498; OmpA\_nmem.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA\_membrane; 1.  
DR PRINTS; P01021; OMPADOMAIN.  
DR ProDom; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OMPA; FALSE NEG.  
KW Direct protein sequencing; Fimbria; Outer membrane; Porin; Signal;  
KW Transmembrane.  
FT SIGNAL 1 21  
FT CHAIN 22 359 Outer membrane protein P5.  
FT DISULFID 332 344 By similarity.  
FT DOMAIN 278 322 OmpA-like.  
SQ SEQUENCE 359 AA; 38340 MW; 576B1C59B4818C37 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 359;  
Best Local Similarity 77.8%; Pred. No. 47;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9  
:|:|:|:|:|:  
Db 138 RSDYKFYED 146

## RESULT 9

ID Q6XE78 PRELIMINARY; PRT; 404 AA.  
AC Q6XE78;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Agul.  
GN Name=agul;  
OS uncultured bacterium.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22894188; PubMed=14532085;  
RA Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E.,  
RA Streit W.R.;  
RT "Prospecting for novel biocatalysts in a soil metagenome.";  
RL Appl. Environ. Microbiol. 69:6235-6242(2003).  
DR EMBL; AY236223; AAP70366.1; -.  
SQ SEQUENCE 404 AA; 46151 MW; 58CEB2949EB94AE4 CRC64;

Query Match 59.7%; Score 43; DB 2; Length 404;  
Best Local Similarity 72.7%; Pred. No. 53;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRI 12  
:|:|:|:|:|:  
Db 76 TDPKFYNNKRI 86

## RESULT 10

AAP70366  
ID AAP70366 PRELIMINARY; PRT; 404 AA.  
AC AAP70366;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
DE Agul.  
GN AGUL.

OS Gamma-proteobacterium Hot 75m4.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E.,  
RA Streit W.R.;  
RT "Prospecting for Novel Biocatalysts in a Soil Metagenome.";  
RL Appl. Environ. Microbiol. 69:6235-6242(2003).  
DR EMBL; AY236223; AAP70366.1; -.  
SQ SEQUENCE 404 AA; 46151 MW; 58CEB2949EB94AE4 CRC64;

Query Match 59.7%; Score 43; DB 2; Length 404;  
Best Local Similarity 72.7%; Pred. No. 53;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRI 12  
:|:|:|:|:|:  
Db 76 TDPKFYNNKRI 86

## RESULT 11

Q6XE55 PRELIMINARY; PRT; 451 AA.  
ID Q6XE55  
AC Q6XE55;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Agul (Agul).  
GN Name=agul; Synonyms=aguC;  
OS uncultured bacterium.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22894188; PubMed=14532085;  
RA Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E.,  
RA Streit W.R.;  
RT "Prospecting for novel biocatalysts in a soil metagenome.";  
RL Appl. Environ. Microbiol. 69:6235-6242(2003).  
DR EMBL; AY236225; AAP70389.1; -.  
DR EMBL; AY212800; AAP49314.1; -.  
SQ SEQUENCE 451 AA; 51461 MW; DB9CE0C8628521E9 CRC64;

Query Match 59.7%; Score 43; DB 2; Length 451;  
Best Local Similarity 72.7%; Pred. No. 59;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRI 12  
:|:|:|:|:|:  
Db 123 TDPKFYNNKRI 133

## RESULT 12

AAP49314  
ID AAP49314 PRELIMINARY; PRT; 451 AA.  
AC AAP49314;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
DE Agul.  
GN AGUL.  
OS Gamma-proteobacterium Hot 75m4.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E.,  
RA Streit W.R.;  
RT "Prospecting for Novel Biocatalysts in a Soil Metagenome.";  
RL Appl. Environ. Microbiol. 69:6235-6242(2003).  
DR EMBL; AY212800; AAP49314.1; -.  
SQ SEQUENCE 451 AA; 51461 MW; DB9CE0C8628521E9 CRC64;



Query Match 59.7%; Score 43; DB 2; Length 451;  
 Best Local Similarity 72.7%; Pred. No. 59;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SDYKFDNKR 12  
 Db 123 TDPKFDNKR 133

RESULT 13  
 AAP70389 PRELIMINARY; PRT; 451 AA.  
 ID AAP70389  
 AC AAP70389;  
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE AGUI.  
 GN AGUI.  
 OS Gamma-proteobacterium Hot 75m4.  
 OC Bacteria; environmental samples.  
 OX NCBI\_TaxID=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E.,  
 RA Streit W.R.;  
 RT "Prospecting for Novel Biocatalysts in a Soil Metagenome."  
 RL Appl. Environ. Microbiol. 69:6235-6242(2003).  
 DR EMBL; AY236225; AAP70389.1; -.  
 SQ SEQUENCE 451 AA; 51461 MW; D59CEOC8628521B9 CRC64;

Query Match 59.7%; Score 43; DB 2; Length 451;  
 Best Local Similarity 72.7%; Pred. No. 59;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SDYKFDNKR 12  
 Db 123 TDPKFDNKR 133

RESULT 14  
 Y659 PASMU STANDARD; PRT; 1905 AA.  
 ID Y659 PASMU  
 AC Q9CWX1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hypothetical UPF0192 protein PM0659 precursor.  
 GN OrderedLocustNames=PM0659;  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Em70."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- SIMILARITY: Belongs to the UPF0192 family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE006102; AAK02743.1; -.  
 DR InterPro; IPR011047; Quin\_alc\_DH\_like.  
 DR InterPro; IPR004112; Succ\_DH\_Flav\_C.

KW Complete proteome; Hypothetical protein; Signal.  
 FT SIGNAL 1 16 Potential.  
 FT CHAIN 17 1905 Hypothetical UPF0192 protein PM0659.  
 SQ SEQUENCE 1905 AA; 214427 MW; FED71CE8D61F7C78 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 1905;  
 Best Local Similarity 60.0%; Pred. No. 2.7e+02;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SDYKFDNKR 11  
 Db 805 ADYKFDNKR 814

RESULT 15  
 Q8IIIG1 PRELIMINARY; PRT; 2616 AA.  
 ID Q8IIIG1  
 AC Q8IIIG1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=PF11\_0213;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Bertman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Pauleen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Partea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014839; AAN35797.1; -.  
 DR InterPro; IPR011591; Botulinum.  
 DR InterPro; IPR000717; PCI.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR ProDom; PD01963; Botulinum; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2616 AA; 313007 MW; C8934D1C5E188B42 CRC64;

Query Match 59.7%; Score 43; DB 2; Length 2616;  
 Best Local Similarity 63.6%; Pred. No. 3.9e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 13  
 Db 1933 EYKFDNKR 1943

RESULT 16  
 Q6MS53 PRELIMINARY; PRT; 124 AA.  
 ID Q6MS53  
 AC Q6MS53;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical prolipoprotein.  
 GN OrderedLocustNames=MSC\_0927;  
 OS Mycoplasma mycoides (subsp. mycoides SC).  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=44101;  
 RN [1]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN=PGI;
RX PubMed14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J.,
RA Johansson K.-E., Petersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PGIT, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842645; CAE77537.1; -.
KW Complete proteome; Hypothetical protein; Lipoprotein.
SQ SEQUENCE 124 AA; 14612 MW; 5B45D717CA3070CA CRC64;

Query Match 58.3%; Score 42; DB 2; Length 124;
Best Local Similarity 58.3%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
Db 32 KOELKFYDNKNI 43

RESULT 17
CAE77537 PRELIMINARY; PRT; 124 AA.
AC CAE77537;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical prolipoprotein.
GN MSC_0927.
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RP SEQUENCE FROM N.A.
RC STRAIN=PGI;
RX PubMed14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J.,
RA Johansson K.-E., Petersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PGIT, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842645; CAE77537.1; -.
KW Hypothetical protein; Lipoprotein.
SQ SEQUENCE 124 AA; 14612 MW; 5B45D717CA3070CA CRC64;

Query Match 58.3%; Score 42; DB 2; Length 124;
Best Local Similarity 58.3%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12
Db 32 KOELKFYDNKNI 43

RESULT 18
Q8IK12 PRELIMINARY; PRT; 248 AA.
AC Q8IK12;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF10_0028;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

```

```

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014829; AAN35226.1; -.
DR HSPF; P09651; IHA1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; RRM_1; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
KW Hypothetical protein.
SQ SEQUENCE 248 AA; 30051 MW; 2120EC2D2E14A688 CRC64;

Query Match 58.3%; Score 42; DB 2; Length 248;
Best Local Similarity 53.8%; Pred. No. 47;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 13
Db 172 RNDYRNYDRSID 184

RESULT 19
Q9UWV0 PRELIMINARY; PRT; 286 AA.
AC Q9UWV0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein ORF-c22_006 (Hypothetical protein SS00545).
GN Names=ORF-c22_006; OrderedLocusNames=SS00545;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2;
RX MEDLINE=20165948; PubMed=10701121;
RA Charlebois R.L., Singh R.K., Chan-Weiher C.C.Y., Allard G., Chow C.,
RA Confalonieri F., Curtis B., Duguet M., Erasuo G., Raguy D.,
RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA St Jean A., Van Der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA Ragan M.A., Sensen C.W.;
RT "Gene content and organization of a 281-kbp contig from the genome of
RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";
RL Genome 43:116-136(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; Y18930; CAB57754.1; -.
DR NCBI_TaxID=36329;
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

```

```

Query Match      58.3%; Score 42; DB 2; Length 286;
Best Local Similarity 53.8%; Pred. No. 55;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13
   :|||:|:|:|:|
Db 53 KSYNYFYDSSKID 65

RESULT 20
ID Q7RNK0 PRELIMINARY; PRT; 360 AA.
AC Q7RNK0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Putative yir1 protein.
GN Name=PY01693;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Augioli S.V., Suh B.B., Kooij T.W., Perteu M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
RA Shallon S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
CC -1- EMBL/GenBank/DDbj whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000456; EAA21053.1; -.
DR InterPro; IPR006477; Yir_bir_cir.
DR Pfam; PF06022; Cir_Bir_Yir; 1.
DR TIGRFAMs; TIGR01590; Yir-bir-cir_Pla; 1.
SQ SEQUENCE 360 AA; 42467 MW; 8E2D069FD49F199A CRC64;

Query Match      58.3%; Score 42; DB 2; Length 360;
Best Local Similarity 58.3%; Pred. No. 70;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRID 13
   :|||:|:|:|:|
Db 41 SDYDFHNGKLE 52

RESULT 21
Q8A4H6 PRELIMINARY; PRT; 517 AA.
ID Q8A4H6;
AC Q8A4H6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative alpha-1,6-mannanase.
GN OrderedLocusNames=BT2623;
OS Bacteroides thetaioamicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;

```

```

RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaioamicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; AE016936; AA077730.1; -.
DR InterPro; IPR005198; Glyco_hydro_76.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF03663; Glyco_hydro_76; 1.
KW Complete proteome.
SQ SEQUENCE 517 AA; 59079 MW; 87D3392046B0592A CRC64;

Query Match      58.3%; Score 42; DB 2; Length 517;
Best Local Similarity 63.6%; Pred. No. 1e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRIR 12
   :|||:|:|:|:|
Db 346 NDYLFYDNVRL 356

RESULT 22
Q7VXJ8 PRELIMINARY; PRT; 870 AA.
ID Q7VXJ8;
AC Q7VXJ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative exported protein.
GN OrderedLocusNames=Bp1760;
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toham I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640416; CAB42047.1; -.
KW Complete proteome.
SQ SEQUENCE 870 AA; 98017 MW; 3DE949DBF06DA4CD CRC64;

Query Match      58.3%; Score 42; DB 2; Length 870;
Best Local Similarity 58.3%; Pred. No. 1.8e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRIR 12
   :|||:|:|:|:|
Db 731 RADYFYNGKRIR 742

RESULT 23
Q7W8Y6 PRELIMINARY; PRT; 870 AA.
ID Q7W8Y6;
AC Q7W8Y6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative exported protein.
GN OrderedLocusNames=BPP1991;
OS Bordetella parapertussis.

```

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12822 / ATCC BAA-587;  
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngi1227;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leath S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 DR EMBL: BX640429; CAE37291.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 870 AA; 97990 MW; FDF49DBF070A4C4 CRC64;  
 Query Match 58.3%; Score 42; DB 2; Length 870;  
 Best Local Similarity 58.3%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 RSDYKFDYDKRI 12  
 :||| :| :|||  
 DB 731 RADYVYNGKRI 742

RESULT 24  
 Q7WKC9 PRELIMINARY; PRT; 870 AA.  
 ID Q7WKC9;  
 AC Q7WKC9;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative exported protein.  
 GN OrderedLocusNames=BB2179;  
 OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngi1227;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leath S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 DR EMBL: BX640443; CAE32675.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 870 AA; 98017 MW; 3DB949DBF06DA4CD CRC64;  
 Query Match 58.3%; Score 42; DB 2; Length 870;  
 Best Local Similarity 58.3%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 RSDYKFDYDKRI 12  
 :||| :| :|||

DB 731 RADYVYNGKRI 742

RESULT 25  
 Q9A9I5 PRELIMINARY; PRT; 903 AA.  
 ID Q9A9I5;  
 AC Q9A9I5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE TonB-dependent receptor; putative.  
 GN OrderedLocusNames=CC0995;  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Poczcka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H.M., Shetty J.,  
 RA Berry K.J., Utterback T.R., Tran K., Wolf A.M., Vamathevan J.J.,  
 RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 DR EMBL: AE005777; AAK22979.1; -;  
 DR PIR: G87372; G87372.  
 DR TIGR: CC0995; -;  
 DR GO: GO:0019867; C:outer membrane; IEA.  
 DR GO: GO:0004872; F:receptor activity; IEA.  
 DR GO: GO:0005215; F:transporter activity; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR000531; TonB receptor.  
 DR InterPro: IPR010104; TonB\_receptorbac.  
 DR Pfam: PF00593; TonB\_dep\_Rec; 1.  
 DR TIGRFAMs: TIGR01782; TonB-Xanth-Caul; 1.  
 DR Complete proteome; Membrane; Outer membrane; Receptor; TonB box.  
 SQ SEQUENCE 903 AA; 99453 MW; 888933577283E7C9 CRC64;  
 Query Match 58.3%; Score 42; DB 2; Length 903;  
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 RSDYKFDYDKRI 12  
 :||| :| :|||  
 DB 267 RADYKFDNNNSI 278

RESULT 26  
 Q8A3D0 PRELIMINARY; PRT; 1021 AA.  
 ID Q8A3D0  
 AC Q8A3D0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative outer membrane protein, probably involved in nutrient binding.  
 GN OrderedLocusNames=BT3024;  
 OS Bacteroides thetaiotaomicron.  
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 OC Bacteroidaceae; Bacteroides.  
 OX NCBI\_TaxID=818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VPI-5482 / ATCC 29148;  
 RX MEDLINE=22550858; PubMed=12663928;  
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,

```

RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
CC 1- SURCELLULAR LOCATION: Outer membrane (By similarity).
DR EMBL; AE016938; LA078130.1; -.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR008969; Carboxypeptidase.
DR Pfam; PF00531; TonB receptor.
DR Complete proteome: Membrane; Outer membrane; Receptor; TonB box.
KW SEQUENCE 1021 AA; 112777 MW; 9E80FA1FBF59FEF CRC64;
SQ SEQUENCE 1021 AA; 112777 MW; 9E80FA1FBF59FEF CRC64;

Query Match 58.3%; Score 42; DB 2; Length 1021;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFDYDK 10
Db 341 KSDYSFEDNR 350

RESULT 27
Q7YYQ2 PRELIMINARY; PRT; 1102 AA.
AC Q7YYQ2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SNF2 helicase, possible.
GN ORFNames=IMB.528;
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RA Bankier A.T., Spriggs H.F., Fartmann B., Konfortov B.A., Madera M.,
RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;
RT "Integrated mapping, chromosomal sequencing and sequence analysis of
RT Cryptosporidium parvum.";
RL Genome Res. 0:0-0(2003).
DR EMBL; BX538352; CAD98428.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAD_box.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001005; Myb DNA_Binding.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00717; SANT; 2.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; UNKNOWN_1.
KW Helicase.
SQ SEQUENCE 1102 AA; 127708 MW; 0997FBF17BL27274 CRC64;

Query Match 58.3%; Score 42; DB 2; Length 1102;
Best Local Similarity 70.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDYDKRI 12
Db 820 DWQFYDKRI 829

RESULT 28

```

```

Q6MTB3
ID Q6MTB3 PRELIMINARY; PRT; 309 AA.
AC Q6MTB3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Mannose-6-phosphate isomerase (EC 5.3.1.8).
GN Name=pmi; OrderedLocNames=MSC_0495;
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG1;
RX PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PG1T, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842643; CAE77123.1; -.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0004476; F:mannose-6-phosphate isomerase activity; IEA.
DR InterPro; IPR001250; Man6P_isomerase1.
DR InterPro; IPR011051; Km1C_like_cupin.
DR Pfam; PF01238; PMI_type1; 1.
DR Complete proteome; Isomerase.
KW SEQUENCE 309 AA; 35928 MW; 0508225F88197B12 CRC64;

Query Match 56.9%; Score 41; DB 2; Length 309;
Best Local Similarity 70.0%; Pred. No. 89;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 YKFDYDKRID 13
Db 189 YRFYDYNRID 198

RESULT 29
CAE77123 PRELIMINARY; PRT; 309 AA.
ID CAE77123
AC CAE77123;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE Mannose-6-phosphate isomerase (EC 5.3.1.8).
GN PMI OR MSC_0495
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG1;
RX PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PG1T, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842643; CAE77123.1; -.
DR Isomerase.
KW SEQUENCE 309 AA; 35928 MW; 0508225F88197B12 CRC64;

Query Match 56.9%; Score 41; DB 2; Length 309;
Best Local Similarity 70.0%; Pred. No. 89;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 YKFDYDKRID 13
Db 189 YRFYDYNRID 198

```

```
RESULT 30
Q7RO14
ID Q7RO14 PRELIMINARY; PRT; 354 AA.
AC Q7RO14;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01113;
OS Plasmodium yoelii yoelii
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perteu M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519 (2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000295; EAA20366.1; -.
KW Hypothetical protein.
SQ SEQUENCE 354 AA; 41675 MW; 700A43802C5EA7F1 CRC64;

Query Match 56.9%; Score 41; DB 2; Length 354;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 KFYDNKRI 12
DQ 223 KFYDNKKI 230

RESULT 31
HGPI_HAEIN
ID HGPI_HAEIN STANDARD; PRT; 1063 AA.
AC P44795;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 1
DE precursor.
GN OrderedLocusNames=HI0635;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G.G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
```

```

QY 3 DYKFDNKRID 13
  ||| | ||: |
  257 DYKYPNQAD 267

Db 122593 NM; EFB8D5CE4247583 CRC64;

RESULT 32
HGPC_HABIN STANDARD; PRT; 1066 AA.
AC Q9X442;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein C precursor.
GN Name=HgpC;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI689 / Serotype B;
RX MEDLINE=99270928; PubMed=10338475;
RA Morton D.J., Whitby P.W., Stull T.L.;
RT "Effect of multiple mutations in the hemoglobin- and hemoglobin-
RT haptoglobin-binding proteins, Hgpb, Hgpb, and HgpC, of Haemophilus
RT influenzae type b.";
RL Infect. Immun. 67:2729-2739(1999).
CC -!- FUNCTION: Acts as a receptor for hemoglobin or the
CC hemoglobin/haptoglobin complex of the human host and is required
CC for heme uptake.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: This protein is subject to phase-variable
CC expression associated with alteration in the length of the CCAA
CC repeat region. This mechanism is called slipped-strand mispairing.
CC Addition or loss of CCAA repeat units would change the reading
CC frame and result in introduction of stop codons downstream of the
CC repeat region. This may be a mechanism of regulation and a way to
CC avoid the immunological response of the host.
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF094574; AAD33112.1; -.
DR InterPro; IPR006970; PT.
DR InterPro; IPR010916; TonB_Box_N.
DR InterPro; IPR010949; TonB_hemlactrns.
DR InterPro; IPR000531; TonB_receptor.
DR InterPro; IPR010917; TonB_recept_C.
DR Pfam; PF04886; PT; 1.
DR TIGRFAMs; TIGR01785; TonB-dep Rec; 1.
DR TIGRFAMs; TIGR01786; TonB-hemlactrns; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Multigene family; Outer membrane; Receptor; Repeat; Signal; TonB box;
TS SIGNAL 1 24 Potential.
FT CHAIN 25 1066 Hemoglobin and hemoglobin-haptoglobin
FT binding protein C.
FT 7 X 4 AA tandem repeats of Q-P-T-N.
FT DOMAIN 26 53
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.

```

Query Match 56.9%; Score 41; DB 1; Length 1066;  
Best Local Similarity 63.6%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DYKFDNKRID 13  
 ||| | ||: |  
 257 DYKYPNQAD 267

Db 122593 NM; EFB8D5CE4247583 CRC64;

RESULT 33  
HGHB\_HABIN STANDARD; PRT; 1067 AA.  
AC Q9KIVL; 2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hemoglobin and hemoglobin-haptoglobin binding protein B precursor  
DE (Hemoglobin binding protein B).  
GN Name=HgbB;  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTHI N182;  
RX MEDLINE=203116037; PubMed=10859226;  
RA Cope L.D., Hrkal Z., Hansen E.J.;  
RT "Detection of phase variation in expression of proteins involved in  
RT hemoglobin and hemoglobin-haptoglobin binding by nontypeable  
RT Haemophilus influenzae.";  
RL Infect. Immun. 68:4092-4101(2000).  
CC -!- FUNCTION: Acts as a receptor for hemoglobin or the  
CC hemoglobin/haptoglobin complex of the human host and is required  
CC for heme uptake.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- MISCELLANEOUS: This protein is subject to phase-variable  
CC expression associated with alteration in the length of the CCAA  
CC repeat region. This mechanism is called slipped-strand mispairing.  
CC Addition or loss of CCAA repeat units would change the reading  
CC frame and result in introduction of stop codons downstream of the  
CC repeat region. This may be a mechanism of regulation and a way to  
CC avoid the immunological response of the host.  
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;  
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF221059; AAF80177.1; -.  
InterPro; IPR006970; PT.  
InterPro; IPR010916; TonB\_Box\_N.  
InterPro; IPR010949; TonB\_hemlactrns.  
InterPro; IPR000531; TonB\_receptor.  
InterPro; IPR010917; TonB\_recept\_C.  
Pfam; PF04886; PT; 1.  
TIGRFAMs; TIGR01785; TonB-dep Rec; 1.  
TIGRFAMs; TIGR01786; TonB-hemlactrns; 1.  
PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Multigene family; Outer membrane; Receptor; Repeat; Signal; TonB box;

```

KW Transprot.      1 24 Potential.
FT SIGNAL          25 1067 Hemoglobin and hemoglobin-haptoglobin
FT CHAIN           26 49 binding protein B.
FT DOMAIN          26 29 6 X 4 AA tandem repeats of Q-P-T-N.
FT REPEAT          30 33 1.
FT REPEAT          34 37 2.
FT REPEAT          38 41 3.
FT REPEAT          42 45 4.
FT REPEAT          46 49 5.
FT SITE            59 66 TonB box.
FT SITE            1050 1067 TonB C-terminal box.
SQ SEQUENCE        1067 AA; 122471 MW; 29D295DC4747632E CRC64;

Query Match      56.9%; Score 41; DB 1; Length 1067;
Best Local Similarity 63.6%; Pred. No. 3.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DYKFDNKRID 13
   |||||
Db 253 DYKIYPNQAD 263

RESULT 34
ID_HGP3_HAEMIN STANDARD; PRT; 1084 AA.
AC P44836;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 3
DE precursor.
GN OrderedLocustNames=HI0712;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD Rd.";
RL Science 269:496-512(1995).
RN [2]

RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
CC Electrophoresis 21:411-429(2000).
CC -1- FUNCTION: Acts as a receptor for hemoglobin or the
CC hemoglobin/haptoglobin complex of the human host and is required
CC for heme uptake (By similarity).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- MISCELLANEOUS: This protein is subject to phase-variable
CC expression associated with alteration in the length of the CCAA
CC repeat region. This mechanism is called slipped-strand mispairing.
CC Addition or loss of CCAA repeat units would change the reading
CC frame and result in introduction of stop codons downstream of the
CC repeat region. This may be a mechanism of regulation and a way to
CC avoid the immunological response of the host (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;

```

```

CC -----
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32754; AAC22369.1; -.
CC PIR; B64088; B64088.
CC TIGR; HI0712; -.
CC InterPro; IPR006970; PT.
CC InterPro; IPR010916; TONB_Box_N.
CC InterPro; IPR010949; TonB_hemLactrns.
CC InterPro; IPR000531; TonB_receptor.
CC InterPro; IPR010917; TonB_recept_C.
CC Pfam; PF04886; PT; 2.
CC DR TIGRPFAMs; TIGR01785; TonB-hemin; 1.
CC DR TIGRPFAMs; TIGR01786; TonB-hemLactrns; 1.
CC DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC KW Complete proteome; Multigene family; Outer membrane; Receptor; Repeat;
KW Signal; TonB box; Transprot.
FT SIGNAL          1 24 Potential.
FT CHAIN           25 1084 Probable hemoglobin and hemoglobin-
FT DOMAIN          26 73 haptoglobin binding protein 3.
FT REPEAT          26 29 12 X 4 AA tandem repeats of Q-P-T-N.
FT REPEAT          30 33 1.
FT REPEAT          34 37 2.
FT REPEAT          38 41 3.
FT REPEAT          42 45 4.
FT REPEAT          46 49 5.
FT REPEAT          50 53 6.
FT REPEAT          54 57 7.
FT REPEAT          58 61 8.
FT REPEAT          62 65 9.
FT REPEAT          66 69 10.
FT REPEAT          70 73 11.
FT SITE            83 90 TonB box.
FT SITE            1067 1084 TonB C-terminal box.
SQ SEQUENCE        1084 AA; 123955 MW; 794DF91E0F53CFD9 CRC64;

Query Match      56.9%; Score 41; DB 1; Length 1084;
Best Local Similarity 63.6%; Pred. No. 3.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DYKFDNKRID 13
   |||||
Db 277 DYKIYPNQAD 287

RESULT 35
Q89T91 PRELIMINARY; PRT; 172 AA.
AC Q89T91;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE B112155 protein.
GN OrderedLocustNames=b112155;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,

```





RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the ompA family.

DR EMBL; AY643798; AAT57680.1; -.  
DR InterPro; IPR006664; Bac\_OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR006890; OMPA\_LIKE.  
DR InterPro; IPR000498; OmpA\_tmem.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA\_membrane; 1.  
DR PRINTS; PR01021; OMPADOMAIN.  
DR PRINTS; PR01022; OUTRMMBRANE.  
DR ProDom; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OMPA; 1.  
SQ SEQUENCE 353 AA; 37964 MW; 1B1399FC2B27DBA CRC64;

Query Match 55.6%; Score 40; DB 2; Length 353;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSDYKFYDN 9  
|||||  
Db 137 RSDYKVDH 145

## RESULT 40

Q6GUB5  
ID Q6GUB5 PRELIMINARY; PRT; 353 AA.  
AC Q6GUB5;  
DT 05-JUL-2004 (TRENBLrel. 27, Created)  
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
DE OmpA.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T931317;  
RA Akridge H., Confer A.W., Dabo S.M.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the ompA family.  
DR EMBL; AY643797; AAT57679.1; -.  
DR InterPro; IPR006664; Bac\_OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR006890; OMPA\_LIKE.  
DR InterPro; IPR000498; OmpA\_tmem.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA\_membrane; 1.  
DR PRINTS; PR01021; OMPADOMAIN.  
DR PRINTS; PR01022; OUTRMMBRANE.  
DR ProDom; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OMPA; 1.  
SQ SEQUENCE 353 AA; 38034 MW; 529CC0E642804220 CRC64;

Query Match 55.6%; Score 40; DB 2; Length 353;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSDYKFYDN 9  
|||||  
Db 137 RSDYKVDH 145

Search completed: November 24, 2004, 09:28:55  
Job time : 61.7717 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:44 ; Search time 10.7391 Seconds  
(without alignments)  
116.473 Million cell updates/sec

Title: US-09-719-379A-3

Perfect score: 72

Sequence: 1 RSDYKFYDNKRID 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 42    | 58.3        | 286    | 2 H90200 | hypothetical prote  |
| 2          | 42    | 58.3        | 903    | 2 G87372 | TonB-dependent rec  |
| 3          | 41    | 56.9        | 953    | 2 B64083 | hemoglobin-binding  |
| 4          | 41    | 56.9        | 1084   | 2 B64088 | hemoglobin-binding  |
| 5          | 40    | 55.6        | 408    | 2 T29857 | probable aspartate  |
| 6          | 40    | 55.6        | 428    | 2 T28007 | hypothetical prote  |
| 7          | 40    | 55.6        | 437    | 1 F64782 | probable transport  |
| 8          | 40    | 55.6        | 463    | 2 D90700 | probable transport  |
| 9          | 40    | 55.6        | 463    | 2 G85550 | probable transport  |
| 10         | 40    | 55.6        | 466    | 2 F98843 | protein ZK809.1 [i  |
| 11         | 40    | 55.6        | 1289   | 2 T30681 | DNA-directed RNA p  |
| 12         | 40    | 55.6        | 1646   | 2 T40198 | hypothetical SPBC3  |
| 13         | 39.5  | 54.9        | 505    | 2 T28276 | ORF MSV115 probabl  |
| 14         | 39.5  | 54.9        | 557    | 2 B28182 | hemolysin B - Serr  |
| 15         | 39    | 54.2        | 264    | 2 G82939 | conserved hypothet  |
| 16         | 39    | 54.2        | 266    | 2 D90131 | hypothetical prote  |
| 17         | 39    | 54.2        | 638    | 2 D69957 | conserved hypothet  |
| 18         | 39    | 54.2        | 740    | 2 T22638 | hypothetical prote  |
| 19         | 39    | 54.2        | 921    | 2 A50332 | conserved hypothet  |
| 20         | 38    | 52.8        | 144    | 2 T28286 | hypothetical prote  |
| 21         | 38    | 52.8        | 159    | 2 G72301 | cationic outer mem  |
| 22         | 38    | 52.8        | 170    | 2 F81038 | probable shikimate  |
| 23         | 38    | 52.8        | 207    | 2 B97320 | S-adenosylmethioni  |
| 24         | 38    | 52.8        | 220    | 2 S73866 | hypothetical methio |
| 25         | 38    | 52.8        | 246    | 1 B55582 | cytochrome-c oxida  |
| 26         | 38    | 52.8        | 256    | 1 S10164 | fumarate reductase  |
| 27         | 38    | 52.8        | 325    | 2 C84151 | hypothetical prote  |
| 28         | 38    | 52.8        | 507    | 2 T50054 | probable transport  |
| 29         | 38    | 52.8        | 595    | 2 A03402 | aspartate-tRNA lig  |

|    |    |      |      |          |                    |
|----|----|------|------|----------|--------------------|
| 30 | 38 | 52.8 | 715  | 2 B41842 | lysine decarboxyla |
| 31 | 38 | 52.8 | 715  | 2 G86108 | lysine decarboxyla |
| 32 | 38 | 52.8 | 715  | 2 A98268 | lysine decarboxyla |
| 33 | 38 | 52.8 | 789  | 2 S28259 | androgen-regulated |
| 34 | 38 | 52.8 | 879  | 2 AC1308 | pyruvate phosphate |
| 35 | 38 | 52.8 | 879  | 2 AC1680 | pyruvate phosphate |
| 36 | 38 | 52.8 | 904  | 2 G90563 | lipoprotein (impor |
| 37 | 38 | 52.8 | 955  | 2 S46551 | beta-agarase - Vib |
| 38 | 38 | 52.8 | 1083 | 1 S53048 | alpha-mannosidase  |
| 39 | 38 | 52.8 | 1196 | 2 T24222 | hypothetical prote |
| 40 | 37 | 51.4 | 218  | 2 H84561 | probable inorganic |
| 41 | 37 | 51.4 | 255  | 2 B70116 | hypothetical prote |
| 42 | 37 | 51.4 | 258  | 2 E71646 | hypothetical prote |
| 43 | 37 | 51.4 | 263  | 2 S13379 | inorganic diphosph |
| 44 | 37 | 51.4 | 286  | 2 G81264 | hypothetical prote |
| 45 | 37 | 51.4 | 287  | 2 S42655 | nodulin-31 precurs |

ALIGNMENTS

RESULT 1

H90200

hypothetical protein SSO0545 [imported] - Sulfolobus solfataricus

C;Species: Sulfolobus solfataricus

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004

C;Accession: H90200

R;She, Q.; Singh, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A;Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139

A;Accession: H90200

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-286 <KUR>

A;Cross-references: UNIPROT:Q9UWV0; GB:AB006641; NID:gl3813707; PIDN:AAK40863.1; GSPDB:G

C;Genetics:

A;Gene: SSO0545

Query Match

Best Local Similarity 58.3%; Score 42; DB 2; Length 286;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qv 1 RSDYKFYDNKRID 13

Db 53 KSYNFFYDSSKID 65

RESULT 2

G87372

TonB-dependent receptor, probable [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: G87372

R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: G87372

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-903 <STO>

A;Cross-references: UNIPROT:Q9A9I5; GB:AB005673; NID:gl3422281; PIDN:AAK22379.1; GSPDB:G

C;Genetics:

A;Gene: CC0995

Query Match

Best Local Similarity 58.3%; Score 42; DB 2; Length 903;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy 1 RSDYKFDNKR 12
|:|||||
Db 267 RADIYKFDNN 278

RESULT 3
B64083
hemoglobin-binding protein homolog HI0635 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 22-Oct-1999
A:Accession: B64083
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64083
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-953 <TIGR>
A:Cross-references: GB:U32746; GB:L42023; NID:g1573626; PIDN:AAC22294.1; PID:g1573631; T
C:Genetics:
A:Start codon: GTG
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
F:1-130/Domain: tonB-dependent receptor amino-terminal homology <TNN>

Query Match 56.9%; Score 41; DB 2; Length 953;
Best Local Similarity 63.6%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 13
|:|||||
Db 147 DYKIYPNQAD 157

RESULT 4
B64088
hemoglobin-binding protein homolog HI0712 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
A:Accession: B64088
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64088
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1084 <TIGR>
A:Cross-references: UNIPROT:P44836; GB:U32754; GB:L42023; NID:g1573711; PIDN:AAC22369.1;
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
F:118-260/Domain: tonB-dependent receptor amino-terminal homology <TNN>

Query Match 56.9%; Score 41; DB 2; Length 1084;
Best Local Similarity 63.6%; Pred. No. 74;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKFDNKR 13
|:|||||
Db 277 DYKIYPNQAD 287

RESULT 5
T29857
probable aspartate transaminase (EC 2.6.1.1) T01C8.5 [similarity] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
```

```
C:Accession: T29857
R:Wohlmann, P.; Hawkins, J.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid T01C8.
A:Reference number: 220699
A:Accession: T29857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-408 <WOH>
A:Cross-references: UNIPROT:Q22067; EMBL:U58726; PIDN:AAB00578.1; GSPDB:GN000028; CESP:T0
A:Experimental source: strain Bristol N2; clone T01C8
C:Genetics:
A:Gene: CESP:T01C8.5
A:Map position: X
A:Introns: 133/2; 257/1; 360/1
C:Superfamily: aspartate aminotransferase
C:Keywords: aminotransferase; pyridoxal phosphate
F:251/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 55.6%; Score 40; DB 2; Length 408;
Best Local Similarity 61.5%; Pred. No. 42;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

Qy 2 SDYKFDNKR 12
|:|||||
Db 149 ADYTFWDYDNKR 161

RESULT 6
T28007
hypothetical protein ZK809.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28007
R:Dobson, R.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z20454
A:Accession: T28007
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-428 <WIL>
A:Cross-references: UNIPROT:Q23602; EMBL:Z68303; PIDN:CAA92641.2; GSPDB:GN000022; CESP:ZK
A:Experimental source: clone ZK809
C:Genetics:
A:Gene: CESP:ZK809.1
A:Map position: 4
A:Introns: 2/2; 192/3; 358/3; 427/1

Query Match 55.6%; Score 40; DB 2; Length 428;
Best Local Similarity 63.6%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFDNKR 12
|:|||||
Db 163 TDYKCYDHNRI 173

RESULT 7
F64782
probable transport protein b0511 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C:Accession: F64782
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64782
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-437 <BLAT>
A:Cross-references: GB:AE000157; GB:U00096; NID:g1786716; PIDN:AAC73613.1; PID:g1786721;
```

A;Experimental source: strain K-12, substrain MG1655  
C;Superfamily: Escherichia coli probable transport protein b0511  
C;Keywords: transmembrane protein

F;15-31/Domain: transmembrane #status predicted <TM1>  
F;82-98/Domain: transmembrane #status predicted <TM2>  
F;105-121/Domain: transmembrane #status predicted <TM3>  
F;143-159/Domain: transmembrane #status predicted <TM4>  
F;179-195/Domain: transmembrane #status predicted <TM5>  
F;221-237/Domain: transmembrane #status predicted <TM6>  
F;259-275/Domain: transmembrane #status predicted <TM7>  
F;304-320/Domain: transmembrane #status predicted <TM8>  
F;328-344/Domain: transmembrane #status predicted <TM9>  
F;375-391/Domain: transmembrane #status predicted <TM10>  
F;403-421/Domain: transmembrane #status predicted <TM11>

Query Match 55.6%; Score 40; DB 1; Length 437;  
Best Local Similarity 85.7%; Pred. No. 45;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:|  
Db 367 DYKYDN 373

# RESULT 8

D90700  
probable transport protein ECs0572 [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic  
A;Reference number: A39629; MUID:21156231; PMID:11258796

A;Accession: D90700

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-463 <HAY>

A;Cross-references: UNIPROT:Q8XCX4; GB:BA000007; PIDN:BA833995.1; PID:gl3360030; GSPDB:G  
A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: ECs0572

C;Superfamily: Escherichia coli probable transport protein b0511

Query Match 55.6%; Score 40; DB 2; Length 463;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:|  
Db 414 DYKYDN 420

# RESULT 9

G85550

probable transport protein Z0665 [similarity] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004

C;Accession: G85550

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dialanta, E.; Potamoudis, K.; Apodaca,

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: G85550

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-463 <STO>

A;Cross-references: UNIPROT:Q8XCX4; GB:AE005174; NID:gl2513405; PIDN:AAG54867.1; GSPDB:G  
A;Experimental source: strain O157:H7, substrain ED1933

C;Genetics:

A;Gene: Z0665

C;Superfamily: Escherichia coli probable transport protein b0511

Query Match 55.6%; Score 40; DB 2; Length 463;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||:|  
Db 414 DYKYDN 420

# RESULT 10

F88843

protein ZK809.1 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: F88843

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A;Reference number: A75000; MUID:99069613; PMID:9831916

A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: F88843

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-466 <STO>

A;Cross-references: UNIPROT:Q23602; GB:chr\_IV; PIDN:CAA92641.1; PID:g3881799; GSPDB:GN00

C;Genetics:

A;Gene: ZK809.1

A;Map position: 4

Query Match 55.6%; Score 40; DB 2; Length 466;  
Best Local Similarity 63.6%; Pred. No. 48;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRI 12  
:|:|:|:|:|  
Db 163 TDYKCYDHNRI 173

# RESULT 11

T30681

DNA-directed RNA polymerase (EC 2.7.7.6) 146K chain - Mollusca contagiosum virus 1

C;Species: Mollusca contagiosum virus 1

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004

C;Accession: T30681

R;Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Dairai, G.; Moss, B.

Science 273, 813-816, 1996

A;Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re

A;Reference number: Z20876; MUID:96325459; PMID:8670425

A;Accession: T30681

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1289 <SEN>

A;Cross-references: UNIPROT:Q98246; EMBL:U60315; PIDN:AAC55207.1

C;Genetics:

A;Note: MC079R

C;Superfamily: vaccinia virus DNA-directed RNA polymerase 147K chain

C;Keywords: nucleotidyltransferase, transcription

Query Match 55.6%; Score 40; DB 2; Length 1289;  
Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRI 12  
:|:|:|:|:|  
Db 331 RNEVKFYFNKRL 342

# RESULT 12

T40198

hypothetical SPBC31E1.01c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004  
C:Accession: T40198; T40630  
R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: Z21912  
A:Accession: T40198  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1595 <OLI>  
A:Cross-references: UNIPROT:Q94549; EMBL:AL049190; PIDN:CAB39135.1; GSPDB:GN00067; SPDB:  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: Z21941  
A:Accession: T40630  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1596-1646 <LYN>  
A:Cross-references: EMBL:AL034563; PIDN:CAA22538.1; GSPDB:GN00067; SPDB:SPBC660.18c  
C:Genetics:  
A:Gene: SPDB:SPBC31E1.01c  
A:Map position: 2

Query Match 55.6%; Score 40; DB 2; Length 1646;  
Best Local Similarity 54.5%; Pred. NO. 1.6e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DYKFYDNKRID 13  
|||:::|  
Db 1033 DYNFFENRGID 1043

RESULT 13  
T38276  
ORF MSV115 probable vaccinia G5R homolog - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28276  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28276  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-505 <AFO>  
A:Cross-references: UNIPROT:Q9YVX7; EMBL:AF063866; NID:g4049647; PIDN:AAC97659.1; PID:g4  
C:Genetics:  
A:Note: MSV115  
C:Superfamily: vaccinia virus probable 49.8K protein

Query Match 54.9%; Score 39.5; DB 2; Length 505;  
Best Local Similarity 50.0%; Pred. NO. 63;  
Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

Qy 1 RSDYKFYD---NKRI 12  
:|||||  
Db 472 KCDYKFYDHLNKL I 486

RESULT 14  
B28182  
hemolysin B - Serratia marcescens  
C:Species: Serratia marcescens  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 09-Jul-2004  
C:Accession: B28182  
R:Poole, K.; Schiebel, E.; Braun, V.  
J. Bacteriol. 170, 3177-3188, 1988  
A:Title: Molecular characterization of the hemolysin determinant of Serratia marcescens.  
A:Reference number: A28182; MUID:88257037; PMID:3290200  
A:Accession: B28182  
A:Molecule type: DNA

A:Residues: 1-557 <POO>  
A:Cross-references: UNIPROT:P15321; GB:M22618; NID:g340726; PIDN:AAA50322.1; PID:G556419  
C:Keywords: transmembrane protein

Query Match 54.9%; Score 39.5; DB 2; Length 557;  
Best Local Similarity 64.3%; Pred. NO. 69;  
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 1 RSDYKFY-DNKRID 13  
|||||:::|  
Db 322 RSDYVFYRHDQID 335

RESULT 15  
G82939  
conserved hypothetical UU061 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: G82939  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min  
A:Reference number: A82870  
A:Accession: G82939  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-264 <GLA>  
A:Cross-references: GB:AE002105; GB:AF222894; NID:g6899003; PIDN:AAF30466.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: UU061  
A:Genetic code: SGC3

Query Match 54.2%; Score 39; DB 2; Length 264;  
Best Local Similarity 85.7%; Pred. NO. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKFYDN 9  
|||||  
Db 133 DYKFYEN 139

RESULT 16  
D90131  
hypothetical protein orf266 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D90131  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; MUID:11323671; PMID:11323671  
A:Accession: D90131  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-266 <DOU>  
A:Cross-references: UNIPROT:Q98S40; GB:AF083031; NID:gl3794366; PIDN:AAK39743.1; GSPDB:G  
C:Genetics:  
A:Gene: orf266  
A:Map position: 3  
A:Genome: nucleomorph  
C:Keywords: nucleomorph

Query Match 54.2%; Score 39; DB 2; Length 266;  
Best Local Similarity 85.7%; Pred. NO. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 YKFYDNK 10  
|||::|  
Db 103 KYFYDNK 109

RESULT 17  
D69957  
conserved hypothetical protein yqgs - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: D69957  
R:Kunet, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertel  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle  
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, K.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: D69957  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-638 <KUN>  
A:Cross-references: UNIPROT:P54496; GB:Z99116; GB:AL009126; NID:G2634723; PIDN:CAB14415.  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yqgs  
C:Superfamily: Bacillus subtilis probable anion-binding protein yf1e

Query Match 54.2%; Score 39; DB 2; Length 638;  
Best Local Similarity 66.7%; Pred. NO. 96;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 5 KFYDNKRID 13  
Db 615 RFYDQKRLD 623

RESULT 18  
T22638  
hypothetical protein F54C9.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T22638  
R:Sims, M.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: Z19591  
A:Accession: T22638  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-740 <WTL>  
A:Cross-references: UNIPROT:Q20760; EMBL:Z49967; PIDN:CAA90257.1; GSPDB:GN00020; CESP:FS  
A:Experimental source: clone F54C9  
C:Genetics:  
A:Gene: CESP:F54C9.9  
A:Map position: 2  
A:Introns: 28/3; 46/1; 140/2; 169/2; 257/2; 295/3; 409/3; 481/1; 651/3; 705/3

Query Match 54.2%; Score 39; DB 2; Length 740;  
Best Local Similarity 58.3%; Pred. NO. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 SDYKFYDNKRID 13  
Db 413 SDYEDYDNLVD 424

RESULT 19  
AE0332  
conserved hypothetical protein YPO2725 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AE0332  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AE0332  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-921 <KUR>  
A:Cross-references: UNIPROT:Q8ZD67; GB:AL590842; PIDN:CAC92964.1; PID:gl5980703; GSPDB:G  
C:Genetics:  
A:Gene: YPO2725

Query Match 54.2%; Score 39; DB 2; Length 921;  
Best Local Similarity 54.5%; Pred. NO. 1.4e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 SDYKFYDNKR 12  
Db 126 SDPKYQKNV 136

RESULT 20  
T28286  
hypothetical protein 125 - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28286  
R:Afonso, C.L.; Tullman, E.R.; Lu, Z.; Oms, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28286  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-144 <AFO>  
A:Cross-references: UNIPROT:Q9YVW7; EMBL:AF063866; NID:G4049647; PIDN:AAC97798.1; PID:G4  
C:Genetics:  
A:Note: MSV125  
C:Superfamily: Melanoplus sanguinipes entomopoxvirus hypothetical protein 125

Query Match 52.8%; Score 38; DB 2; Length 144;  
Best Local Similarity 50.0%; Pred. NO. 33;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 RSDYKFYDNKR 12  
Db 82 KDYYIYNKKI 93

RESULT 21  
G72301  
cationic outer membrane protein - Thermotoga maritima (strain MS88)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: G72301  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: G72301  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <ARN>  
A:Cross-references: UNIPROT:Q9X0E3; GB:AE001765; GB:AE000512; NID:G4981586; PIDN:AAD3613  
A:Experimental source: strain MSB8  
C:Genetics:

A:Gene: TM1053

Query Match 52.8%; Score 38; DB 2; Length 159;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RSDYKFYDNK 10  
: ||| |||  
Db 54 KRDYSFYQNK 63

#### RESULT 22

F81038  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: F81038; H81984  
R:Petelin, H.; Saunders, N.J.; Heideberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.B.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizsa, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: F81038  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-170 <TET>  
A:Cross-references: UNIPROT:Q9JQV1; GB:AE002531; GB:AE002098; NID:g7227065; PIDN:AAF4214  
A:Experimental source: serogroup B, strain MC58  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20225556; PMID:10761919  
A:Accession: H81984  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-170 <PAR>  
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83937.1; PID:g3737937  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: aroK; NMB1813; NMA0648  
C:Superfamily: shikimate kinase, shikimate kinase homology  
C:Keywords: phosphotransferase

Query Match 52.8%; Score 38; DB 2; Length 170;  
Best Local Similarity 66.7%; Pred. No. 39;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSDYKFYDN 9  
: ||| |||  
Db 28 RLDYRFYDS 36

#### RESULT 23

B97320  
S-adenosylmethionine-dependent methyltransferase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 12-Jul-2004  
C:Accession: B97320  
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97320  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-207 <KUP>  
A:Cross-references: UNIPROT:Q97DQ3; GB:AE001437; PIDN:AAK81349.1; PID:g15026507; GSPDB:C  
A:Experimental source: Clostridium acetobutylicum ATCC8294  
C:Genetics:

A:Gene: CAC3419  
C:Superfamily: spore germination protein C2

Query Match 52.8%; Score 38; DB 2; Length 207;  
Best Local Similarity 46.2%; Pred. No. 48;  
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSDYKFYDNKRID 13  
: ||| |||  
Db 167 KGDYKLYSKKEIE 179

#### RESULT 24

S73866  
hypothetical protein H10\_orf220L - Mycoplasma pneumoniae (strain ATCC 29342)  
C:Species: Mycoplasma pneumoniae  
A:Variety: ATCC 29342  
C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: S73866  
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.  
Nucleic Acids Res. 24, 4420-4449, 1996  
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A:Reference number: S73327; MUID:97105885; PMID:8948633  
A:Accession: S73866  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-220 <HIM>  
A:Cross-references: UNIPROT:P75482; EMBL:AE000053; GB:U00089; NID:g1674236; PIDN:AAB9618  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C:Genetics:  
A:Genetic code: SGC3  
A:Start codon: TTG

Query Match 52.8%; Score 38; DB 2; Length 220;  
Best Local Similarity 54.5%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 DYKFYDNKRID 13  
: ||| |||  
Db 16 DYRFYHEKFD 26

#### RESULT 25

B55582  
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Azorhizobium caulinodans  
N:Alternate names: cb-type cytochrome-c oxidase 28K chain; cytochrome b410; fixO protein  
C:Species: Azorhizobium caulinodans  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: B55582; S42230  
R:Mandon, K.; Kaminski, P.A.; Elmerich, C.  
J. Bacteriol. 176, 2560-2568, 1994  
A:Title: Functional analysis of the fixNOOP region of Azorhizobium caulinodans.  
A:Reference number: A55582; MUID:94222833; PMID:8169204  
A:Accession: B55582  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-246 <MAN1>  
A:Cross-references: UNIPROT:Q43943; GB:X74410; NID:g456310; PIDN:CAA52430.1; PID:g456312  
R:Mandon, K.; Kaminski, P.A.; Muegel, C.; Desnoues, N.; Dreyfus, B.; Elmerich, C.  
FEMS Microbiol. Lett. 114, 185-190, 1993  
A:Title: Role of the fixGHI region of Azorhizobium caulinodans in free-living and symbio  
A:Reference number: S42229; MUID:94109675; PMID:8282187  
A:Accession: S42230  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-246 <MAN2>

A:Cross-references: EMBL:X74410; NID:g456310; PIDN:CAA52430.1; PID:g456312  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993  
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain  
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory  
Query Match 52.8%; Score 38; DB 1; Length 246;  
Best Local Similarity 75.0%; Pred. No. 56;



Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDYDK 10  
|:|:|:|:|  
Db 235 DFKLYDNK 242

RESULT 26  
S10164  
fumarate reductase (EC 1.3.99.1) cytochrome b component - Wolinella succinogenes  
N;Alternate names: fumarate reductase chain C  
C;Species: Wolinella succinogenes  
C;Date: 21-Nov-1993 #sequence\_revision 19-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S10164; A44954  
R;Koertner, C.; Lauterbach, F.; Tripiet, D.; Unden, G.; Kroeger, A.  
Mol. Microbiol. 4, 855-860, 1990  
A;Title: Wolinella succinogenes fumarate reductase contains a dihaem cytochrome b.  
A;Reference number: S10164; MUID:90355847; PMID:2388563  
A;Accession: S10164  
A;Status: preliminary; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-256 <KOB>  
A;Cross-references: UNIPROT:P17413; EMBL:X51509; NID:G48511; PIDN:CAA35874.1; PID:G48512  
R;Lauterbach, F.; Koertner, C.; Albracht, S.P.J.; Unden, G.; Kroeger, A.  
Arch. Microbiol. 154, 386-393, 1990  
A;Title: The fumarate reductase operon of Wolinella succinogenes. Sequence and expression  
A;Reference number: A44954; MUID:91058386; PMID:2244791  
A;Accession: A44954  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 205-256 <LAU>  
A;Cross-references: GB:X51509  
C;Genetics:  
A;Gene: frdC  
C;Complex: part of an enzyme complex containing a heterotrimer (flavoprotein, iron-sulfur  
C;Function:  
A;Description: catalyzes the oxidation of succinate to fumarate and transfers its reduced  
f the complex together with the iron sulfur subunit  
C;Superfamily: fumarate reductase, cytochrome b subunit  
C;Keywords: oxidoreductase; transmembrane protein; tricarboxylic acid cycle

Query Match 52.8%; Score 38; DB 1; Length 256;  
Best Local Similarity 66.7%; Pred. No. 58;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFDYDKR 11  
|:|:|:|:|  
Db 244 DYKFDYDKR 252

RESULT 27  
C84151  
hypothetical protein BH4011 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C;Accession: C84151  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: C84151  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-325 <STO>  
A;Cross-references: UNIPROT:Q9K5S7; GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BA8077  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH4011

Query Match 52.8%; Score 38; DB 2; Length 325;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKEYDN 9  
|:|:|:|:|  
Db 104 DYKEYDD 110

RESULT 28  
T50054  
probable transporter [imported] - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
C;Accession: T50054  
R;Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A;Reference number: 225031  
A;Accession: T50054  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-507 <HUN>  
A;Cross-references: UNIPROT:Q9US37; EMBL:AL133521; PIDN:CAB63540.1; GSPDB:GN00066; SPDB:  
A;Experimental source: strain 972h(-); cosmid c1039  
C;Genetics:  
A;Gene: SPDB:SPAC1039.04  
A;Map position: 1

Query Match 52.8%; Score 38; DB 2; Length 507;  
Best Local Similarity 58.3%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKFYDNKRID 13  
|:|:|:|:|  
Db 466 SVYMFDFNRRRD 477

RESULT 29  
AD3402  
aspartate-tRNA ligase (EC 6.1.1.12) [imported] - Brucella melitensis (strain 16M)  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C;Accession: AD3402  
R;DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujter, C.; Los, T.; Ivanova,  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A;Reference number: AD3402; PMID:1175668  
A;Accession: AD3402  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-595 <KUR>  
A;Cross-references: UNIPROT:Q8YGF7; GB:AE008917; PIDN:AAL52383.1; PID:G17983182; GSPDB:G  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI1202  
A;Map position: 1  
C;Superfamily: lysine-tRNA ligase  
C;Keywords: ligase

Query Match 52.8%; Score 38; DB 2; Length 595;  
Best Local Similarity 53.3%; Pred. No. 1.3e+02;  
Matches 8; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

Qy 3 DYKFY----DNKRID 13  
|:|:|:|:|  
Db 437 DFFPFEWDEDNKKID 451

RESULT 30  
B41842  
lysine decarboxylase (EC 4.1.1.18) cada - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C;Accession: B41842; S09407; S56360; A41968; B65223  
R;Meng, S.Y.; Bennett, G.N.  
J. Bacteriol. 174, 2659-2669, 1992



```

Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKPYDNKRID 13
   :||| :||| :|||
Db 399 NDYPCGNGKQVD 410

RESULT 34
AC1308
pyruvate phosphate dikinase homolog lml867 [imported] - Listeria monocytogenes (strain
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1308
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669.
A:Accession: AC1308
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-879 <GLA>
A:Cross-references: UNIPROT:Q8Y633; GB:NC_003210; PIDN:CAC99945.1; PID:gl6411321; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lml867
C:Superfamily: pyruvate, phosphate dikinase

Query Match 52.8%; Score 38; DB 2; Length 879;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKPYDNKRKI 12
   :||| :||| :|||
Db 561 RTHMFFDEKRI 572

RESULT 35
AC1680
pyruvate phosphate dikinase homolog lin1981 [imported] - Listeria innocua (strain Clip1
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1680
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1680
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-879 <GLA>
A:Cross-references: UNIPROT:Q92AE1; GB:AL592022; PIDN:CAC97211.1; PID:gl6414482; GSPDB:C
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin1981
C:Superfamily: pyruvate, phosphate dikinase

Query Match 52.8%; Score 38; DB 2; Length 879;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKPYDNKRKI 12
   :||| :||| :|||
Db 561 RTHMFFDEKRI 572

```

```

RESULT 36
G90563
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: G90563
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: G90563
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-904 <KUR>
A:Cross-references: UNIPROT:Q98QF1; GB:AL445566; PID:g1409829; PIDN:CAC13588.1; GSPDB:G
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU 4150
A:Genetic code: SGC3

Query Match 52.8%; Score 38; DB 2; Length 904;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKPYDNKRKI 12
   :||| :||| :|||
Db 235 RKEYKFDENRKI 246

RESULT 37
S46651
beta-agarase - Vibrio sp. (strain JT0107)
C:Species: Vibrio sp.
A:Variety: strain JT0107
C:Date: 14-Feb-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C:Accession: S46651; S45296
R:Sugano, Y.; Matsumoto, T.; Noma, M.
submitted to the EMBL Data Library, October 1993
A:Reference number: S46651
A:Accession: S46651
A:Molecule type: DNA
A:Residues: 1-955 <SUG>
A:Cross-references: UNIPROT:P48840; EMBL:D21202; NID:g498302; PIDN:BAA04744.1; PID:g55312
A:Experimental source: strain JT0107
R:Sugano, Y.; Matsumoto, T.; Noma, M.
Biochim. Biophys. Acta 1218, 105-108, 1994
A:Title: Sequence analysis of the agab gene encoding a new beta-agarase from Vibrio sp.
A:Reference number: S45296; MUID:94250684; PMID:8193156
A:Accession: S45296
A:Molecule type: DNA
A:Residues: 497-934 <SUW>
A:Cross-references: EMBL:D21202
A:Experimental source: strain JT0107
C:Genetics:
A:Gene: agab
C:Superfamily: Vibrio agarase

Query Match 52.8%; Score 38; DB 2; Length 955;
Best Local Similarity 54.5%; Pred. No. 2.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKPYDNKRKI 12
   :||| :||| :|||
Db 624 ADPMFYDNKKV 634

RESULT 38
S53048
alpha-mannosidase (EC 3.2.1.24) - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G1861; protein YGL156w
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004

```

C;Accession: S60420; A33511; S64172; S53048  
R;James, C.M.; Indge, K.J.; Oliver, S.G.  
Yeast 11, 1413-1419, 1995  
A;Title: DNA sequence analysis of a 35 kb segment from *Saccharomyces cerevisiae* chromosome  
A;Reference number: S60417; MUID:96158061; PMID:8585324  
A;Accession: S60420  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1083 <JAM1>  
A;Cross-references: UNIPROT:P22855; EMBL:Z48618; NID:G728690; PIDN:CAA88536.1; PID:G7286  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995  
R;Oshihisa, T.; Anraku, Y.  
Biochem. Biophys. Res. Commun. 163, 908-915, 1989  
A;Title: Nucleotide sequence of AMS1, the structure gene of vacuolar alpha-mannosidase c  
A;Reference number: A33511; MUID:89392009; PMID:2675832  
A;Accession: A33511  
A;Molecule type: DNA  
A;Residues: 1-785, 'L', 787-797, 'A', 799-1083 <YOS>  
A;Cross-references: GB:M29146; NID:G171056; PIDN:AAA34423.1; PID:G171058; GB:M27809  
A;Note: the authors translated the codon CTT for residue 786 as Val, and GCT for residue  
R;James, C.M.; Indge, K.J.; Oliver, S.G.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64165  
A;Accession: S64172  
A;Molecule type: DNA  
A;Residues: 1-1083 <JAM2>  
A;Cross-references: EMBL:Z72678; NID:G1322745; PIDN:CAA96868.1; PID:G1322746; GSPDB:GN00  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:AMS1; MIPS:YGL156w  
A;Cross-references: SGD:S0003124; MIPS:YGL156w  
A;Map position: 7L  
C;Superfamily: Saccharomyces alpha-mannosidase  
C;Keywords: blocked amino end; glycosidase; hydrolase; yeast vacuole

Query Match 52.8%; Score 38; DB 1; Length 1083;  
Best Local Similarity 87.5%; Pred. NO. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KFYDNKRI 12  
|||||  
Db 40 KFYDKKRI 47

RESULT 39  
T24222  
hypotheical protein R13H4.1 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T24222  
R;Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19858  
A;Accession: T24222  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1196 <WIL>  
A;Cross-references: UNIPROT:P90948; EMBL:Z81579; PIDN:CAB04654.1; GSPDB:GN000023; CESP:R1  
A;Experimental source: clone R13H4  
C;Genetics:  
A;Gene: CESP:R13H4.1  
A;Map position: 5  
A;Introns: 42/3; 70/1; 109/2; 173/1; 235/3; 281/2; 572/2; 632/3; 725/1; 857/1; 1068/2; 1

Query Match 52.8%; Score 38; DB 2; Length 1196;  
Best Local Similarity 66.7%; Pred. NO. 2.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DYKFYDNKR 11  
|:|:|:|  
Db 388 DFKFYENGR 396

RESULT 40  
H84561  
probable inorganic pyrophosphatase [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: H84561  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: H84561  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-218 <STO>  
A;Cross-references: UNIPROT:P21216; GB:AE002093; NID:G4309743; PIDN:AADI5513.1; GSPDB:GN  
C;Genetics:  
A;Gene: At2g18230  
A;Map position: 2  
C;Superfamily: inorganic pyrophosphatase

Query Match 51.4%; Score 37; DB 2; Length 218;  
Best Local Similarity 54.5%; Pred. NO. 74;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DYKFYDNKRID 13  
|||:|:|:|  
Db 177 DYKKNENKVD 187

Search completed: November 24, 2004, 09:30:17  
Job time : 12.7391 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:07:49 ; Search time 82.8152 Seconds  
(without alignments)  
82.302 Million cell updates/sec

Title: US-09-719-379A-1  
Perfect score: 105  
Sequence: 1 RSDYKFEAANGTRDHKKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_23Sep04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID      | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| 1          | 105   | 100.0       | 19     | 3 AAY79959 | Aay79959 Non-typea |
| 2          | 105   | 100.0       | 19     | 3 AAY79987 | Aay79987 Non-typea |
| 3          | 101   | 96.2        | 19     | 3 AAY79960 | Aay79960 Non-typea |
| 4          | 100   | 95.2        | 19     | 3 AAY79961 | Aay79961 Non-typea |
| 5          | 99    | 94.3        | 19     | 3 AAY79982 | Aay79982 Non-typea |
| 6          | 99    | 94.3        | 19     | 3 AAY79991 | Aay79991 Non-typea |
| 7          | 99    | 94.3        | 19     | 3 AAY79955 | Aay79955 Non-typea |
| 8          | 99    | 94.3        | 19     | 4 AAB47439 | Aab47439 LBI(f) co |
| 9          | 99    | 94.3        | 20     | 3 AAB20881 | Aab20881 LBIgr1 pe |
| 10         | 99    | 94.3        | 28     | 4 AAB47443 | Aab47443 Entire 3r |
| 11         | 99    | 94.3        | 40     | 2 AAW67581 | Aaw67581 Synthetic |
| 12         | 99    | 94.3        | 40     | 3 AAY79986 | Aay79986 Measles v |
| 13         | 99    | 94.3        | 40     | 6 ADA35172 | Ada35172 Chimeric  |
| 14         | 99    | 94.3        | 40     | 7 ADC89661 | Adc89661 H. influe |
| 15         | 99    | 94.3        | 359    | 2 AAR62294 | Aar62294 Plasmid L |
| 16         | 99    | 94.3        | 464    | 3 AAY79993 | Aay79993 Non-typea |
| 17         | 96    | 91.4        | 19     | 3 AAY79957 | Aay79957 Non-typea |
| 18         | 95    | 90.5        | 19     | 3 AAY79963 | Aay79963 Non-typea |
| 19         | 94    | 89.5        | 19     | 3 AAY79958 | Aay79958 Non-typea |
| 20         | 94    | 89.5        | 19     | 3 AAY79956 | Aay79956 Non-typea |
| 21         | 93    | 88.6        | 18     | 2 AAW67572 | Aaw67572 Non-typab |
| 22         | 93    | 88.6        | 18     | 6 ADA25163 | Ada25163 H. influe |
| 23         | 91    | 86.7        | 19     | 3 AAY79967 | Aay79967 Non-typea |
| 24         | 91    | 86.7        | 19     | 3 AAY79968 | Aay79968 Non-typea |
| 25         | 90    | 85.7        | 19     | 3 AAY79973 | Aay79973 Non-typea |

|    |    |      |     |            |                    |
|----|----|------|-----|------------|--------------------|
| 26 | 89 | 84.8 | 19  | 3 AAY79970 | Aay79970 Non-typea |
| 27 | 89 | 84.8 | 19  | 3 AAY79966 | Aay79966 Non-typea |
| 28 | 88 | 83.8 | 19  | 3 AAY79962 | Aay79962 Non-typea |
| 29 | 88 | 83.8 | 19  | 3 AAY79965 | Aay79965 Non-typea |
| 30 | 86 | 81.9 | 19  | 3 AAY79971 | Aay79971 Non-typea |
| 31 | 85 | 81.0 | 19  | 3 AAY79992 | Aay79992 Non-typea |
| 32 | 85 | 81.0 | 19  | 3 AAY79964 | Aay79964 Non-typea |
| 33 | 84 | 80.0 | 19  | 3 AAY79969 | Aay79969 Non-typea |
| 34 | 84 | 80.0 | 338 | 2 AAR85450 | Aar85450 Nontypabl |
| 35 | 83 | 79.0 | 18  | 7 ADC89652 | Adc89652 H. influe |
| 36 | 80 | 76.2 | 19  | 3 AAY79972 | Aay79972 Non-typea |
| 37 | 48 | 45.7 | 311 | 3 AAG45896 | Aag45896 Arabidops |
| 38 | 48 | 45.7 | 343 | 3 AAG34578 | Aag34578 Arabidops |
| 39 | 48 | 45.7 | 361 | 3 AAG20945 | Aag20945 Arabidops |
| 40 | 48 | 45.7 | 361 | 3 AAG45883 | Aag45883 Arabidops |
| 41 | 48 | 45.7 | 361 | 3 AAG24458 | Aag24458 Arabidops |
| 42 | 48 | 45.7 | 378 | 3 AAG20944 | Aag20944 Arabidops |
| 43 | 48 | 45.7 | 378 | 3 AAG24457 | Aag24457 Arabidops |
| 44 | 48 | 45.7 | 378 | 3 AAG45882 | Aag45882 Arabidops |
| 45 | 48 | 45.7 | 414 | 3 AAG45881 | Aag45881 Arabidops |

## ALIGNMENTS

RESULT 1  
AAY79959  
ID AAY79959 standard; peptide; 19 AA.  
XX AC AAY79959;  
XX DT 15-MAY-2000 (first entry)  
XX DE Non-typeable H. influenzae group 1 LBI(f) peptide N10567RM.  
XX KW Vaccine; non-typeable Haemophilus influenzae; nchi; infection;  
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbriin protein;  
XX KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;  
XX KW conjunctivitis; lower respiratory tract infection.  
XX OS Haemophilus influenzae.  
XX PN WO9964067-A2.  
XX PD 16-DEC-1999.  
XX PF 28-MAY-1999; 99WO-US011980.  
XX PR 11-JUN-1998; 98GB-00012613.  
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.  
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
XX WPI; 2000-116457/10.  
XX DR Novel antigenic P5-like fimbriin subunit peptides used in vaccines against  
XX PT Haemophilus influenza.  
XX PS Example 1; Page 29; 68pp; English.  
XX CC The present invention describes antigenic P5-like fimbriin subunit  
XX CC peptides (LBI(f) peptides) of P5-like fimbriin proteins from various  
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
XX CC prevention, and treatment of Haemophilus influenzae infections, such as  
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
XX CC infection. The peptides may also be used in vaccines against H.  
XX CC influenzae. Antibodies and probes from the present invention can be used  
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
XX CC the present invention



PN WO9964067-A2.  
 XX 16-DEC-1999.  
 XX 28-MAY-1999; 99WO-US011980.  
 XX 11-JUN-1998; 98GB-00012613.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX Example 1; Page 29; 68pp; English.  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 19 AA;  
 SQ Query Match 95.2%; Score 100; DB 3; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 2.7e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFEANGTRDHKKG 19  
 DB 1 RSDYKFEANGTRDHKKG 19  
 RESULT 5  
 AAY79982  
 ID AAY79982 standard; peptide; 19 AA.  
 XX AC AAY79982;  
 XX 15-MAY-2000 (first entry)  
 DE Non-typeable H. influenzae group 1 Lb1(f) peptide N1128.  
 XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX Haemophilus influenzae.  
 OS WO9964067-A2.  
 PN 16-DEC-1999.  
 XX 28-MAY-1999; 99WO-US011980.  
 XX 11-JUN-1998; 98GB-00012613.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX Example 1; Page 29; 68pp; English.  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 19 AA;  
 SQ Query Match 94.3%; Score 99; DB 3; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 4.1e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFEANGTRDHKKG 19  
 DB 1 RSDYKFEANGTRDHKKG 19  
 RESULT 6  
 AAY79991  
 ID AAY79991 standard; peptide; 19 AA.  
 XX AC AAY79991;  
 XX 15-MAY-2000 (first entry)  
 DE Non-typeable H. influenzae 1128 Group 1 type peptide.  
 XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX Haemophilus influenzae.  
 OS WO9964067-A2.  
 PN 16-DEC-1999.  
 XX 28-MAY-1999; 99WO-US011980.  
 XX 11-JUN-1998; 98GB-00012613.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX Claim 11; Page 46; 68pp; English.  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention

CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 19 AA;

Query Match 94.3%; Score 99; DB 3; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 4.1e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 |||||  
 Db 1 RSDYKFYEDANGTRDHKKG 19

## RESULT 7

AAV79955  
 ID AAV79955 standard; peptide; 19 AA.

AC AAV79955;

DT 15-MAY-2000 (first entry)

DE Non-typeable H. influenzae group 1 LB1(f) peptide N1128.

KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

OS Haemophilus influenzae.

XX WO9964067-A2.

XX 16-DEC-1999.

PF 28-MAY-1999; 99WO-US011980.

XX 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenzae.

XX Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention

XX Sequence 19 AA;

Query Match 94.3%; Score 99; DB 3; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 4.1e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 |||||  
 Db 1 RSDYKFYEDANGTRDHKKG 19

## RESULT 8

AA47439  
 ID AAB47439 standard; peptide; 19 AA.

XX AAB47439;

XX 31-OCT-2001 (first entry)

DE LB1(f) containing peptide from strain nH1-1128 (Group 1 type).

XX surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; nH1; LB1(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.

OS Haemophilus influenzae.

XX WO200161013-A1.

XX 23-AUG-2001.

PF 13-FEB-2001; 2001WO-EP001556.

XX 15-FEB-2000; 2000GB-00003502.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PI Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX WPI; 2001-522599/57.

XX Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.

XX Claim 1; Page 26; 29pp; English.

XX The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these  
 CC peptides contain an LB1(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain nH1128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection

XX Sequence 19 AA;

Query Match 94.3%; Score 99; DB 4; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 4.1e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 |||||  
 Db 1 RSDYKFYEDANGTRDHKKG 19

## RESULT 9

AA20881  
 ID AAB20881 standard; peptide; 20 AA.

XX AAB20881;

XX 03-JAN-2001 (first entry)

DE LB1gr1 peptide SEQ ID NO.19.

XX Immunoglobulin E; IgE; immunogenic; immunogen; Protein D; carrier;



KW prostate cancer; Haemophilus influenzae; vaccine; infectious disease;  
 KW malaria; cytostatic; anti-allergic; nontropic; neuroprotective;  
 KW protozoacide; Alzheimer's disease; allergy.  
 XX  
 OS Synthetic.  
 XX  
 PH Key Location/Qualifiers  
 FT Modified-site 1  
 FT Modified-site /note= "acetylated"  
 FT Modified-site 20  
 FT Modified-site /note= "amidated"  
 XX  
 XX WO2000050077-A1.  
 XX  
 XX 31-AUG-2000.  
 XX  
 XX 22-FEB-2000; 2000WO-EP001457.  
 XX  
 XX 25-FEB-1999; 99GB-00004405.  
 PR 25-FEB-1999; 99GB-00004408.  
 PR 25-FEB-1999; 99GB-00004412.  
 PR 13-AUG-1999; 99GB-00019260.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 XX Coeste M, Lobet Y, Van-Mechelen MP, Verriest C;  
 XX WPI; 2000-572040/53.  
 XX  
 XX Immunogens and vaccine comprising the immunogen useful for preventing and  
 PT treating infectious diseases e.g. malaria and chronic disease e.g.  
 PT cancer, comprises peptide and carrier from protein D of influenzae.  
 XX  
 XX Example 14; Page 34; 53pp; English.  
 PS  
 XX The present invention describes an immunogen (I) comprising a peptide  
 CC (1a) and a carrier (1b) derived from protein D of Haemophilus influenzae  
 CC or its fragment. Also described are: (1) a vaccine comprising (I), and an  
 CC excipient; (2) preparation of (I), comprising conjugating a peptide to  
 CC protein D or its fragment; and (3) preparation of a vaccine of (1),  
 CC comprising formulating (I) with an excipient. (I) has cytostatic,  
 CC anti-allergic, nontropic, neuroprotective and protozoacide activities. (I)  
 CC and the vaccine are useful for the manufacture of a medicament for  
 CC preventing and treating infectious diseases such as malaria or chronic  
 CC disease such as cancer, Alzheimer's disease or allergy in a patient.  
 CC Unlike prior art immunogens, (I) induces high levels of antipeptide  
 CC immune responses while inducing a moderate humoral response against the  
 CC carrier. The present sequence represents an L81grl peptide which was  
 CC coupled through an additional C-terminal cysteine via maleimide to  
 CC protein D in an example from the present invention  
 XX  
 SQ Sequence 20 AA;  
 Query Match 94.3%; Score 99; DB 3; Length 20;  
 Best Local Similarity 94.7%; Pred. No. 4.4e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFYEAAANGTRDHKKG 19  
 DB 1 RSDYKFYEDANGTRDHKKG 19  
 DE  
 RESULT 10  
 AAB47443  
 ID AAB47443 standard; peptide; 28 AA.  
 XX  
 AC AAB47443;  
 XX  
 XX 31-OCT-2001 (first entry)  
 DT  
 XX Entire 3rd loop from strain nH1-1128 (Group 1 type).  
 DE  
 XX surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW

KW non-typeable H. influenzae; nH1; LB1(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO200161013-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 XX 13-FEB-2001; 2001WO-EP001556.  
 XX  
 XX 15-FEB-2000; 2000GB-00003502.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 XX Berthet FJ, Denoel P, Poolman J, Thonnard J;  
 XX WPI; 2001-522599/57.  
 XX  
 XX Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.  
 XX  
 XX Claim 2; Page 26; 29pp; English.  
 XX  
 XX The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these  
 CC peptides contain an LB1(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain nH1128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX  
 SQ Sequence 28 AA;  
 Query Match 94.3%; Score 99; DB 4; Length 28;  
 Best Local Similarity 94.7%; Pred. No. 6.9e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFYEAAANGTRDHKKG 19  
 DB 3 RSDYKFYEDANGTRDHKKG 21  
 DE  
 RESULT 11  
 AAW67581  
 ID AAW67581 standard; peptide; 40 AA.  
 XX  
 AC AAW67581;  
 XX  
 XX 02-MAR-1999 (first entry)  
 DT  
 XX Synthetic chimera fimbrin/T-cell epitope peptide LB1.  
 DE  
 XX Chimeric; non-typable Haemophilus influenzae; fimbrin; T-cell epitope;  
 KW immunogenic composition; immune response.  
 KW  
 XX Synthetic.  
 OS  
 XX US5843464-A.  
 PN  
 XX 01-DEC-1998.  
 PD  
 XX 02-JUN-1995; 95US-00460502.  
 PF  
 XX 02-JUN-1995; 95US-00460502.  
 PR

XX (OHIS ) UNIV OHIO STATE.  
 XX Kaumaya PTP, Bakaletz LO;  
 XX WPI; 1999-044514/04.  
 DR Synthetic chimeric fimbria peptide - useful for vaccination against non-  
 XX typable Haemophilus influenzae.  
 XX Claim 4; Col 4; 16pp; English.  
 XX The invention relates to the manufacture of a synthetic chimeric peptide  
 CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via  
 CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is  
 CC used in immunogenic compositions which induce an immune response against  
 CC non-typable Haemophilus influenzae. This sequence represents an example  
 CC of a chimeric fimbria/T-cell epitope peptide and is designated LBI. The  
 CC peptide comprises a 19 amino acid sequence corresponding to amino acids  
 CC 117-135 of the fimbria protein, the linker sequence and amino acid 288-  
 CC 302 of the measles virus fusion protein (a T-cell epitope)  
 XX  
 XX Sequence 40 AA;  
 SQ

Query Match 94.3%; Score 99; DB 2; Length 40;  
 Best Local Similarity 94.7%; Pred. No. 9.9e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 Db 1 RSDYKFVEDANGTRDHKKG 19

RESULT 12  
 AAY79986  
 ID AAY79986 standard; peptide; 40 AA.  
 XX  
 AC AAY79986;  
 XX  
 XX 15-MAY-2000 (first entry)  
 XX  
 XX Measles virus fusion protein T-cell promiscuous epitope.  
 DE  
 XX Vaccine; non-typable Haemophilus influenzae; nTHi; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Measles virus.  
 OS Synthetic.  
 XX WO9964067-A2.  
 XX  
 XX 16-DEC-1999.  
 XX  
 XX 28-MAY-1999; 99WO-US011980.  
 XX  
 XX 11-JUN-1998; 98GB-00012613.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 XX WPI; 2000-116457/10.  
 XX  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 FT Haemophilus influenzae.  
 XX  
 XX Example 4; Page 38; 68pp; English.  
 PS  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LBI(f) peptides) of P5-like fimbria proteins from various

CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 40 AA;  
 SQ

Query Match 94.3%; Score 99; DB 3; Length 40;  
 Best Local Similarity 94.7%; Pred. No. 9.9e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 Db 1 RSDYKFVEDANGTRDHKKG 19

RESULT 13  
 ADA25172  
 ID ADA25172 standard; peptide; 40 AA.  
 XX  
 AC ADA25172;  
 XX  
 XX 20-NOV-2003 (first entry)  
 XX  
 XX Chimeric fimbria peptide LBI.  
 DE  
 XX fimbria; non-typable Haemophilus influenzae; NTHi infection;  
 KW otitis media.  
 XX  
 OS Chimeric.  
 OS Synthetic.  
 OS Haemophilus influenzae.  
 OS Measles virus.  
 XX US6436405-B1.  
 XX  
 XX 20-AUG-2002.  
 XX  
 XX 04-SEP-1998; 98US-00148711.  
 XX  
 XX 02-JUN-1995; 95US-00460502.  
 XX  
 XX (OHIS ) UNIV OHIO STATE.  
 XX  
 XX Bakaletz LO, Kaumaya PTP;  
 XX  
 XX WPI; 2003-615247/58.  
 XX  
 XX Synthetic chimeric fimbria peptide, useful for treating Haemophilus  
 FT influenzae infections.  
 XX  
 XX Claim 10; Col 4; 16pp; English.  
 XX  
 XX The invention relates to a synthetic chimeric fimbria peptide. The  
 CC peptide is useful for treating a non-typable Haemophilus influenzae  
 CC (NTHi) infection and otitis media. The synthetic peptides do not require  
 CC tedious purification techniques. The present sequence represents the  
 CC amino acid sequence of the chimeric fimbria peptide LBI.  
 XX  
 XX Sequence 40 AA;  
 SQ

Query Match 94.3%; Score 99; DB 6; Length 40;  
 Best Local Similarity 94.7%; Pred. No. 9.9e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHKKG 19  
 Db 1 RSDYKFVEDANGTRDHKKG 19

RESULT 14  
 ADC89661  
 ID ADC89661 standard; peptide; 40 AA.  
 XX AC ADC89661;  
 XX AC ADC89661;  
 DT 01-JAN-2004 (first entry)  
 XX DE H. influenzae fimbrin peptide/T cell epitope chimaera LBI.  
 XX KW Fimbrin; T cell epitope; vaccine; otitis media; auditory;  
 XX KW antiinflammatory; LBI.  
 OS Chimeric.  
 OS Haemophilus influenzae.  
 OS Measles virus.  
 XX US2003113344-A1.  
 XX 19-JUN-2003.  
 XX 19-AUG-2002; 2002US-00223711.  
 XX 04-SEP-1998; 98US-00148711.  
 XX (BAKA/) BAKALETZ L O.  
 XX (KAUM/) KAUMAYA P T P.  
 XX Bakaletz LO, Kaumaya PTP;  
 XX WPI; 2003-810881/76.  
 XX Novel synthetic chimeric fimbrin peptide LBI or LB2 comprising a first  
 PT peptide unit, T cell epitope as second peptide unit and third linker  
 PT peptide unit, useful for preventing or reducing severity of otitis media.  
 XX Claim 8; SEQ ID NO 10; 15pp; English.  
 XX The invention relates to a synthetic chimaeric fimbrin peptide LBI or LB2  
 CC comprises a first peptide unit derived from H. influenzae fimbrin, a  
 CC second peptide unit containing a T cell epitope and a third linker  
 CC peptide which connects the first peptide to the second. The chimaeric  
 CC peptide is useful for inducing an immune response in animals against non-  
 CC typable Haemophilus influenzae (NTHi) and for preventing or reducing the  
 CC adherence of NTHi to host cells thereby preventing or reducing the  
 CC severity of otitis media. The present sequence is an H. influenzae  
 CC fimbrin peptide/measles virus T cell epitope chimaeric peptide of the  
 CC invention, LBI.  
 XX SQ Sequence 40 AA;  
 Query Match 94.3%; Score 99; DB 7; Length 40;  
 Best Local Similarity 94.7%; Pred. No. 9.9e-10;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFYEAAANGTRDHKKG 19  
 DB 1 RSDYKFVEDANGTRDHKKG 19  
 RESULT 15  
 AAR66294  
 ID AAR66294 standard; protein; 359 AA.  
 XX AC AAR66294;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 09-AUG-1995 (first entry)  
 XX DE Non-typable Haemophilus influenza (NTHi) fimbrin protein.  
 XX

KW Fimbrin protein; vaccine; otitis media.  
 OS Haemophilus influenzae.  
 XX Key Location/Qualifiers  
 XX Region 22..33  
 FT /label= amino terminus  
 FT Peptide 234..249  
 FT /label= internal CNBr fragment  
 XX WO9426304-A1.  
 XX 24-NOV-1994.  
 XX 12-MAY-1994; 94WO-US005477.  
 XX 18-MAY-1993; 93US-00065442.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Kolattukudy PE, Bakaletz LO, Sirakova T;  
 XX WPI; 1995-006359/01.  
 XX N-PSDB; AAQ78916.  
 XX Vaccine comprising non-typable Haemophilus influenza fimbrin protein -  
 PT useful in studying, preventing or reducing the severity of otitis media,  
 PT also fimbrin protein and DNA.  
 XX Disclosure; Fig 5; 45pp; English.  
 XX The fimbrin proteins from 15 randomly selected type b and non-typable  
 CC clinical isolates of Haemophilus influenzae share common epitopes. Thus  
 CC fimbrin isolated from non-typable Haemophilus influenzae 1128 strain is  
 CC a particularly suitable immunogen to protect against the different non-  
 CC typable H7. influenzae that cause otitis media. Fimbrin protein is  
 CC produced by culturing a transformed microbial host, pref. E.coli  
 CC Sporodoptera frugiperda or a mucosal pathogen. Fimbrin protein (FP)  
 CC produced by this process is claimed. The FP protein migrates in  
 CC polyacrylamide gels to a posn. equiv. to a mol. wt. of 25.5 kD or 37.5  
 CC kD. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003  
 CC to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
 XX SQ Sequence 359 AA;  
 Query Match 94.3%; Score 99; DB 2; Length 359;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RSDYKFYEAAANGTRDHKKG 19  
 DB 138 RSDYKFVEDANGTRDHKKG 156  
 RESULT 16  
 AAY79993  
 ID AAY79993 standard; protein; 464 AA.  
 XX AC AAY79993;  
 XX 15-MAY-2000 (first entry)  
 XX Plasmid LPD-LBI-III protein sequence.  
 XX Vaccine; non-typable Haemophilus influenzae; nTHi; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;  
 KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX Haemophilus influenzae.  
 OS Synthetic.  
 XX WO9964067-A2.  
 PN

```

XX PD 16-DEC-1999.
XX XX
XX PF 28-MAY-1999; 99WO-US011980.
XX XX
XX PR 11-JUN-1998; 98GB-00012613.
XX XX
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX XX
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX XX
XX DR WPI; 2000-116457/10.
XX DR N-ESDB; AAZ91252.
XX XX
XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PS Claim 14; Fig 5; 68pp; English.
XX XX
XX CC The present invention describes antigenic P5-like fimbria subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AA79955 to AA79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention
XX SQ Sequence 464 AA;

Query Match 94.3%; Score 99; DB 3; Length 464;
Best Local Similarity 94.7%; Pred. No. 1.8e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHKKG 19
Db 416 RSDYKFVEDANGTRDHKKG 434

RESULT 17
AA79957
ID AA79957 standard; peptide; 19 AA.
XX XX
XX AC AA79957;
XX XX
XX DT 15-MAY-2000 (first entry)
XX XX
XX DE Non-typeable H. influenzae group 1 LB1(f) peptide N1234MEE.
XX XX
XX KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX XX
XX OS Haemophilus influenzae.
XX XX
XX FN WO9964067-A2.
XX XX
XX PD 16-DEC-1999.
XX XX
XX PF 28-MAY-1999; 99WO-US011980.
XX XX
XX PR 11-JUN-1998; 98GB-00012613.
XX XX
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX XX
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX XX
XX DR WPI; 2000-116457/10.

```

```

XX XX
XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX XX
XX PS Example 1; Page 29; 68pp; English.
XX XX
XX CC The present invention describes antigenic P5-like fimbria subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AA79955 to AA79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention
XX SQ Sequence 19 AA;

Query Match 91.4%; Score 96; DB 3; Length 19;
Best Local Similarity 89.5%; Pred. No. 1.4e-09;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHKKG 19
Db 1 RSDYKFYDDANGTRDHKKG 19

RESULT 18
AA79963
ID AA79963 standard; peptide; 19 AA.
XX XX
XX AC AA79963;
XX XX
XX DT 15-MAY-2000 (first entry)
XX XX
XX DE Non-typeable H. influenzae group 1 LB1(f) peptide N1848NP.
XX XX
XX KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX XX
XX OS Haemophilus influenzae.
XX XX
XX FN WO9964067-A2.
XX XX
XX PD 16-DEC-1999.
XX XX
XX PF 28-MAY-1999; 99WO-US011980.
XX XX
XX PR 11-JUN-1998; 98GB-00012613.
XX XX
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX XX
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX XX
XX DR WPI; 2000-116457/10.
XX XX
XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX XX
XX PS Example 1; Page 29; 68pp; English.
XX XX
XX CC The present invention describes antigenic P5-like fimbria subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AA79955 to AA79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention

```

CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 19 AA;

Query Match 90.5%; Score 95; DB 3; Length 19;  
 Best Local Similarity 94.4%; Pred. No. 2.1e-09;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFEAAANGTRDHKK 18  
 |||||  
 Db 1 RSDYKFEVANGTRDHKK 18  
 |||||

## RESULT 19

AAZ9958  
 ID AAY79958 standard; peptide; 19 AA.

XX AC AAY79958;

XX 15-MAY-2000 (first entry)

XX Non-typeable H. influenzae group 1 LB1(f) peptide N90100RM.

XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX Haemophilus influenzae.

XX WO9964067-A2.

XX 16-DEC-1999.

XX 28-MAY-1999; 99WO-US011980.

XX 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenzae.

XX Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention

XX Sequence 19 AA;

Query Match 89.5%; Score 94; DB 3; Length 19;  
 Best Local Similarity 89.5%; Pred. No. 3.1e-09;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFEAAANGTRDHKK 19  
 |||||  
 Db 1 RSDYKFEDENGTRDHKK 19  
 |||||

## RESULT 20

AAZ9956  
 ID AAY79956 standard; peptide; 19 AA.

XX AC AAY79956;

XX 15-MAY-2000 (first entry)

XX Non-typeable H. influenzae group 1 LB1(f) peptide N152NP.

XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX Haemophilus influenzae.

XX WO9964067-A2.

XX 16-DEC-1999.

XX 28-MAY-1999; 99WO-US011980.

XX 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenzae.

XX Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention

XX Sequence 19 AA;

Query Match 89.5%; Score 94; DB 3; Length 19;  
 Best Local Similarity 89.5%; Pred. No. 3.1e-09;  
 Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFEAAANGTRDHKK 19  
 |||||  
 Db 1 RSDYKFEADGTRDHKK 19  
 |||||

## RESULT 21

AAZ9957  
 ID AAY67572 standard; peptide; 18 AA.

XX AC AAY67572;

XX 02-MAR-1999 (first entry)

XX Non-typeable H. influenzae fimbria peptide #1.

XX Chimeric; non-typeable Haemophilus influenzae; fimbria; T-cell epitope;  
 KW immunogenic composition; immune response.

XX Haemophilus influenzae.

XX US5843464-A.  
 XX 01-DEC-1998.  
 XX 02-JUN-1995; 95US-00460502.  
 XX 02-JUN-1995; 95US-00460502.  
 XX (OHIS ) UNIV OHIO STATE.  
 XX Kaumaya PTP, Bakaletz LO;  
 XX WPI; 1999-044514/04.  
 XX Synthetic chimeric fimbriin peptide - useful for vaccination against non-  
 XX typable Haemophilus influenzae.  
 XX Claim 1; Col 3; 16pp; English.  
 XX The invention relates to the manufacture of a synthetic chimeric peptide  
 XX comprising a non-typable Haemophilus influenzae fimbriin peptide fused via  
 XX a linker peptide to a T-cell epitope peptide. The chimeric peptide is  
 XX used in immunogenic compositions which induce an immune response against  
 XX non-typable Haemophilus influenzae. This sequence represents an example  
 XX of a H. influenzae fimbriin peptide used to generate the chimeric peptide  
 XX Sequence 18 AA;  
 Query Match 88.6%; Score 93; DB 2; Length 18;  
 Best Local Similarity 94.4%; Pred. No. 4.4e-09;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 RSDYKFYEANGTRDHKK 18  
 Db 1 RSDYKFYEDANGTRDHKK 18  
 RESULT 22  
 ADA25163  
 ID ADA25163 standard; peptide; 18 AA.  
 XX AC ADA25163;  
 XX 20-NOV-2003 (first entry)  
 XX H. influenzae fimbriin subunit peptide #1.  
 XX fimbriin; non-typable Haemophilus influenzae; NTHI infection;  
 XX otitis media.  
 XX Haemophilus influenzae.  
 XX US6436405-B1.  
 XX 20-AUG-2002.  
 XX 04-SEP-1998; 98US-00148711.  
 XX 02-JUN-1995; 95US-00460502.  
 XX (OHIS ) UNIV OHIO STATE.  
 XX Bakaletz LO, Kaumaya PTP;  
 XX WPI; 2003-615247/58.  
 XX Synthetic chimeric fimbriin peptide, useful for treating Haemophilus  
 XX influenzae infections.  
 XX Claim 1; Col 3; 16pp; English.  
 XX The invention relates to a synthetic chimeric fimbriin peptide. The

CC peptide is useful for treating a non-typable Haemophilus influenzae  
 CC (NTHI) infection and otitis media. The synthetic peptides do not require  
 CC tedious purification techniques. The present sequence represents the  
 CC amino acid sequence of H. influenzae fimbriin subunit peptide #1.  
 XX Sequence 18 AA;  
 Query Match 88.6%; Score 93; DB 6; Length 18;  
 Best Local Similarity 94.4%; Pred. No. 4.4e-09;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 RSDYKFYEANGTRDHKK 18  
 Db 1 RSDYKFYEDANGTRDHKK 18  
 RESULT 23  
 AAY79967  
 ID AAY79967 standard; peptide; 19 AA.  
 XX AC AAY79967;  
 XX 15-MAY-2000 (first entry)  
 XX Non-typable H. influenzae group 1 LB1(f) peptide NTHI-601.  
 XX Vaccine; non-typable Haemophilus influenzae; nTHI; infection;  
 XX chimeric protein; Haemophilus influenzae; P5-like fimbriin protein;  
 XX lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 XX conjunctivitis; lower respiratory tract infection.  
 XX Haemophilus influenzae.  
 XX WO9964067-A2.  
 XX 16-DEC-1999.  
 XX 28-MAY-1999; 99WO-US011980.  
 XX 11-JUN-1998; 98GB-00012613.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbriin subunit peptides used in vaccines against  
 XX Haemophilus influenzae.  
 XX Example 1; Page 29; 68pp; English.  
 XX The present invention describes antigenic P5-like fimbriin subunit  
 XX peptides (LB1(f) peptides) of P5-like fimbriin proteins from various  
 XX Haemophilus influenzae strains. The peptides are used for diagnosis,  
 XX prevention, and treatment of Haemophilus influenzae infections, such as  
 XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 XX infection. The peptides may also be used in vaccines against H  
 XX influenzae. Antibodies and probes from the present invention can be used  
 XX for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 XX AA291201 to AA291252, represent sequences used in the exemplification of  
 XX the present invention  
 XX Sequence 19 AA;  
 Query Match 86.7%; Score 91; DB 3; Length 19;  
 Best Local Similarity 88.9%; Pred. No. 1e-08;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 RSDYKFYEANGTRDHKK 18  
 Db 1 RSDYKFYEVANGTRDHKK 18

## RESULT 24

AAV79968  
ID AAY79968 standard; peptide; 19 AA.  
XX  
AC AAV79968;  
XX  
DT 15-MAY-2000 (first entry)  
XX  
DE Non-typeable H. influenzae group 1 LB1(f) peptide N226NP.  
XX  
DE Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;  
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.  
XX  
OS Haemophilus influenzae.  
XX  
XX WO9964067-A2.  
PN  
XX 16-DEC-1999.  
PD  
XX 28-MAY-1999; 99WO-US011980.  
PF  
XX 11-JUN-1998; 98GB-00012613.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
XX  
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
XX WPI; 2000-116457/10.  
DR  
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
PT Haemophilus influenzae.  
XX  
XX Example 1; Page 29; 68pp; English.  
XX  
XX The present invention describes antigenic P5-like fimbria subunit  
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
CC prevention, and treatment of Haemophilus influenzae infections, such as  
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
CC infection. The peptides may also be used in vaccines against H.  
CC influenzae. Antibodies and probes from the present invention can be used  
CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 19 AA;

Query Match 86.7%; Score 91; DB 3; Length 19;  
Best Local Similarity 88.9%; Pred. No. 1e-08;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFEAANGTRDHRK 18  
|||||  
Db 1 RSDYKFEAANGTRDHRK 18

## RESULT 25

AAV79973  
ID AAY79973 standard; peptide; 19 AA.  
XX  
XX AAY79973;  
AC  
XX 15-MAY-2000 (first entry)  
DT  
XX Non-typeable H. influenzae group 1 LB1(f) peptide NTHI-499.  
DE  
XX Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;  
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
XX

KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.  
XX  
OS Haemophilus influenzae.  
XX  
PN WO9964067-A2.  
XX  
PD 16-DEC-1999.  
XX  
XX 28-MAY-1999; 99WO-US011980.  
PF  
XX 11-JUN-1998; 98GB-00012613.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
XX  
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
XX WPI; 2000-116457/10.  
DR  
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
PT Haemophilus influenzae.  
XX  
XX Example 1; Page 30; 68pp; English.  
XX  
XX The present invention describes antigenic P5-like fimbria subunit  
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
CC prevention, and treatment of Haemophilus influenzae infections, such as  
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
CC infection. The peptides may also be used in vaccines against H.  
CC influenzae. Antibodies and probes from the present invention can be used  
CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 19 AA;

Query Match 85.7%; Score 90; DB 3; Length 19;  
Best Local Similarity 84.2%; Pred. No. 1.6e-08;  
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFEAANGTRDHRK 19  
|||||  
Db 1 RSDYKFEAANGTRDHRK 19

## RESULT 26

AAV79970  
ID AAY79970 standard; peptide; 19 AA.  
XX  
XX AAY79970;

DT 15-MAY-2000 (first entry)

DE Non-typeable H. influenzae group 1 LB1(f) peptide N1657MEE.

XX Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;  
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.  
XX  
OS Haemophilus influenzae.  
XX  
PN WO9964067-A2.  
XX  
PD 16-DEC-1999.  
XX  
XX 28-MAY-1999; 99WO-US011980.  
PF  
XX 11-JUN-1998; 98GB-00012613.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
PA





|  |   |
|--|---|
| Qy   | 1 RSDYKFEAAANGTRDHKK 18<br>   |
| Dd   | 1 RSDYKFNDANGTRDHKK 18<br>  |
| RESULT 29  |   |
| AY79965  |   |
| ID   | AA79965 standard; peptide; 19 AA.   |
| XX   |   |
| AC   | AA79965;  |
| XX   |   |
| DT   | 15-MAY-2000 (first entry)   |
| XX   |   |
| DE   | Non-typeable H. influenzae group 1 LBI(f) peptide NTHI-484.   |
| XX   |   |
| KW   | Vaccine; non-typeable Haemophilus influenzae; nHf; infection;<br>chimeric protein; Haemophilus influenzae; P5-like fimbria protein;<br>lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;<br>conjunctivitis; lower respiratory tract infection.  |
| XX   |   |
| OS   | Haemophilus influenzae.   |
| XX   |   |
| PN   | WO9964067-A2.   |
| XX   |   |
| PD   | 16-DEC-1999.  |
| XX   |   |
| Pf   | 28-MAY-1999; 99WO-USO111980.  |
| XX   |   |
| PR   | 11-JUN-1998; 98GB-00012613.   |
| XX   |   |
| PA   | (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.<br>(OHIS ) UNIV OHIO STATE RES FOUND.   |
| XX   |   |
| PI   | Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  |
| XX   |   |
| DR   | WPI; 2000-116457/10.  |
| XX   |   |
| PT   | Novel antigenic P5-like fimbria subunit peptides used in vaccines against<br>Haemophilus influenza.   |
| XX   |   |
| PS   | Example 1; Page 29; 68pp; English.  |
| CC   | The present invention describes antigenic P5-like fimbria subunit<br>peptides (LBI(f) peptides) of P5-like fimbria proteins from various<br>Haemophilus influenza strains. The peptides are used for diagnosis,<br>prevention, and treatment of Haemophilus influenza infections, such as<br>otitis media, sinusitis, conjunctivitis, or lower respiratory tract<br>infection. The peptides may also be used in vaccines against H.<br>influenzae. Antibodies and probes from the present invention can be used<br>for diagnosis of H. influenzae infection. AA79955 to AA79993, and<br>AAZ91201 to AAZ91252, represent sequences used in the exemplification of<br>the present invention |
| SQ   | Sequence 19 AA;   |
| Query Match 81.9%; Score 86; DB 3; Length 19;<br>Best Local Similarity 83.3%; Pred. No. 7.9e-08;<br>Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0 |   |
| Qy   | 1 RSDYKFEAAANGTRDHKK 18<br>   |
| Dd   | 1 RSDYKFVDPNGTRDHKK 18<br>  |
| RESULT 31  |   |
| AY79992  |   |
| ID   | AA79992 standard; peptide; 19 AA.   |
| XX   |   |
| AC   | AA79992;  |
| XX   |   |
| DT   | 15-MAY-2000 (first entry)   |
| XX   |   |
| DE   | Non-typeable H. influenzae protien P5 Group 1 type peptide.   |
| XX   |   |
| KW   | Vaccine; non-typeable Haemophilus influenzae; nHf; infection;<br>chimeric protein; Haemophilus influenzae; P5-like fimbria protein;<br>lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;<br>conjunctivitis; lower respiratory tract infection.  |
| XX   |   |
| OS   | Haemophilus influenzae.   |
| XX   |   |
| FH   | Key Location/Qualifiers   |
| FT   | Misc-difference 16 /note= "unspecified"   |
| XX   |   |
| PN   | WO9964067-A2.   |

|           |   |
|-----------|---|
| Qy        | 1 RSDYKFEAAANGTRDHKK 18<br>                               |
| Dd        | 1 RSDYKFNDANGTRDHKK 18<br>                                |
| RESULT 30 |   |
| AY79971   |   |
| ID        | AA79971 standard; peptide; 19 AA.                         |
| XX        |   |
| AC        | AA79971;  |
| XX        |   |
| DT        | 15-MAY-2000 (first entry)                                 |
| XX        |   |
| DE        | Non-typeable H. influenzae group 1 LBI(f) peptide N214NP. |

|           |   |
|-----------|---|
| Qy        | 1 RSDYKFEAAANGTRDHKK 19<br>                               |
| Dd        | 1 RSDYKFEDANGTRKHKG 19<br>                                |
| RESULT 30 |   |
| AY79971   |   |
| ID        | AA79971 standard; peptide; 19 AA.                         |
| XX        |   |
| AC        | AA79971;  |
| XX        |   |
| DT        | 15-MAY-2000 (first entry)                                 |
| XX        |   |
| DE        | Non-typeable H. influenzae group 1 LBI(f) peptide N214NP. |

PD 16-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX  
 XX Disclosure; Page 46; 68pp; English.  
 XX  
 XX The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 19 AA;  
 SQ  
 Query Match 81.0%; Score 85; DB 3; Length 19;  
 Best Local Similarity 84.2%; Pred. No. 1.2e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 RSDYKFVEAANGTRDHHKG 19  
 Db 1 RSDYKFVEAANGTRDHHKG 19  
 RESULT 32  
 AAY79964  
 ID AAY79964 standard; peptide; 19 AA.  
 XX  
 AC AAY79964;  
 XX  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Non-typeable H. influenzae group 1 LB1(f) peptide NTHI-567.  
 DE  
 XX Vaccine; non-typeable Haemophilus influenzae; nH; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 XX WO9964067-A2.  
 PN  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
 PT

PT Haemophilus influenza.  
 XX  
 PS Example 1; Page 29; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 19 AA;  
 SQ  
 Query Match 81.0%; Score 85; DB 3; Length 19;  
 Best Local Similarity 84.2%; Pred. No. 1.2e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 RSDYKFVEAANGTRDHHKG 19  
 Db 1 RSDYKFVEAANGTRDHHKG 19  
 RESULT 33  
 AAY79969  
 ID AAY79969 standard; peptide; 19 AA.  
 XX  
 AC AAY79969;  
 XX  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Non-typeable H. influenzae group 1 LB1(f) peptide NTHI-480.  
 DE  
 XX Vaccine; non-typeable Haemophilus influenzae; nH; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 XX WO9964067-A2.  
 PN  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX  
 XX Example 1; Page 29; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention

```

XX SQ Sequence 19 AA;
Query Match      80.0%; Score 84; DB 3; Length 19;
Best Local Similarity 78.9%; Pred. No. 1.8e-07;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFYEANGTRDHKKG 19
    ||||| ||||| ||||| ||||| |||||
Db 1 RSDYKFYEANGTRERKRG 19

RESULT 34
AAR85450
ID AAR85450 standard; protein; 338 AA.
XX AC AAR85450;
XX DT 15-FEB-1996 (first entry)
XX DE Nontypable H. influenzae P5 protein.
XX DS P5 outer membrane protein; vaccine; otitis media; sinusitis;
XX KW chronic pulmonary obstructive disease.
XX OS Haemophilus influenzae.
XX FH Key Location/Qualifiers
FT FT Misc-difference 195
FT /notes="amino acid at position 195 is not identified in
FT the specification"
FT Misc-difference 311
FT /notes="amino acid at position 311 is not identified in
FT the specification"
XX PN EP680765-A1.
XX PD 08-NOV-1995.
XX PF 02-MAY-1995; 95EP-00302996.
XX PR 05-MAY-1994; 94US-00210394.
XX PA (AMCY ) AMERICAN CYANAMID CO.
XX PI Zlotnick GW;
XX DR WPI; 1995-375029/49.
XX PT Purified H.influenzae P5 outer membrane protein - used for preventing
XX reducing susceptibility to or treating H.influenzae infections.
XX PS Disclosure; Page 7-8; 16pp; English.
XX CC Nontypable H. influenzae HI outer membrane protein P5 was isolated by
XX extraction of the outer membrane with detergents and cation-exchange
XX chromatography. P5 (or its peptide fragments) are used in vaccines for
XX prevention of H. influenzae infections implicated in otitis media,
XX sinusitis and chronic pulmonary obstructive disease
XX SQ Sequence 338 AA;
Query Match      80.0%; Score 84; DB 2; Length 338;
Best Local Similarity 84.2%; Pred. No. 5.2e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFYEANGTRDHKKG 19
    ||||| ||||| ||||| ||||| |||||
Db 119 RSDYKFYEAPNSTRDAAKG 137

RESULT 35
ADC89652
ID ADC89652 standard; peptide; 18 AA.
XX AC ADC89652;
XX DT 01-JAN-2004 (first entry)
XX DE H. influenzae fimbriae peptide #1.
XX DS Fimbriae; T cell epitope; vaccine; otitis media; auditory;
XX KW antiinflammatory.
XX OS Haemophilus influenzae.
XX PN US2003113344-A1.
XX PD 19-JUN-2003.
XX PF 19-AUG-2002; 2002US-00223711.
XX PR 04-SEP-1998; 98US-00148711.
XX PA (BAKA/) BAKALETZ L O.
XX (KAUM/) KAUMAYA P T P.
XX PI Bakaletz LO, Kaumaya PTP;
XX DR WPI; 2003-810881/76.
XX DS Novel synthetic chimeric fimbriae peptide LB1 or LB2 comprising a first
XX peptide unit, T cell epitope as second peptide unit and third linker
XX peptide unit, useful for preventing or reducing severity of otitis media.
XX PS Claim 1; SEQ ID NO 1; 15pp; English.
XX CC The invention relates to a synthetic chimeric fimbriae peptide LB1 or LB2
XX comprises a first peptide unit derived from H. influenzae fimbriae, a
XX second peptide unit containing a T cell epitope and a third linker
XX peptide which connects the first peptide to the second. The chimeric
XX peptide is useful for inducing an immune response in animals against non-
XX typable Haemophilus influenzae (NTHi) and for preventing or reducing
XX adherence of NTHi to host cells thereby preventing or reducing the
XX severity of otitis media. The present sequence is an H. influenzae
XX fimbriae peptide for use in the chimeric peptides of the invention.
XX SQ Sequence 18 AA;
Query Match      79.0%; Score 83; DB 7; Length 18;
Best Local Similarity 83.3%; Pred. No. 2.5e-07;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKFYEANGTRDHKK 18
    ||||| ||||| ||||| ||||| |||||
Db 1 RSDYKFVEDLNGTRNHKK 18

RESULT 36
AAV79972
ID AAV79972 standard; peptide; 19 AA.
XX AC AAV79972;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 1 LB1(f) peptide N250NP.
XX KW Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbriae protein;
XX lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.

```

```
XX 16-DEC-1999.
XX PD
XX 28-MAY-1999; 99WO-US011980.
XX PF
XX 11-JUN-1998; 98GB-00012613.
XX PR
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX PA
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX PI
XX WPI; 2000-116457/10.
XX DR
XX Novel antigenic P5-like fimbriin subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PT
XX Example 1; Page 29; 68pp; English.
XX PS
XX The present invention describes antigenic P5-like fimbriin subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbriin proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX CC AA291201 to AA291252, represent sequences used in the exemplification of
XX CC the present invention
XX CC
XX SQ Sequence 19 AA;
Query Match 76.2%; Score 80; DB 3; Length 19;
Best Local Similarity 78.9%; Pred. No. 8.9e-07;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 RSDYKFYEANGTRDHKKG 19
Dy 1 RSDYKFYEANGTRHDKG 19
||||| ||||| ||
RESULT 37
AAG45896
ID AAG45896 standard; protein; 311 AA.
AC AAG45896;
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SRQ ID NO: 57678.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX
XX Arabidopsis thaliana.
XX OS
XX EP1033405-A2.
XX PN
XX
XX 06-SEP-2000.
XX PD
XX
XX 25-FEB-2000; 2000EP-00301439.
XX PF
XX
XX 25-FEB-1999; 99US-0121825P.
XX PR
XX 05-MAR-1999; 99US-0123180P.
XX PR
XX 23-MAR-1999; 99US-0123548P.
XX PR
XX 23-MAR-1999; 99US-0125788P.
XX PR
XX 29-MAR-1999; 99US-0126264P.
XX PR
XX 01-APR-1999; 99US-0126785P.
XX PR
XX 06-APR-1999; 99US-0127462P.
XX PR
XX 08-APR-1999; 99US-0128234P.
XX PR
XX 16-APR-1999; 99US-0128714P.
XX PR
XX 16-APR-1999; 99US-0129845P.
XX PR
19-APR-1999; 99US-0130077P.
XX PR
21-APR-1999; 99US-0130449P.
XX PR
23-APR-1999; 99US-0130510P.
XX PR
23-APR-1999; 99US-0130891P.
XX PR
28-APR-1999; 99US-0131449P.
XX PR
30-APR-1999; 99US-0132048P.
XX PR
04-MAY-1999; 99US-0132407P.
XX PR
05-MAY-1999; 99US-0132485P.
XX PR
06-MAY-1999; 99US-0132486P.
XX PR
06-MAY-1999; 99US-0132487P.
XX PR
07-MAY-1999; 99US-0132863P.
XX PR
11-MAY-1999; 99US-0134256P.
XX PR
14-MAY-1999; 99US-0134218P.
XX PR
14-MAY-1999; 99US-0134219P.
XX PR
14-MAY-1999; 99US-0134221P.
XX PR
14-MAY-1999; 99US-0134370P.
XX PR
18-MAY-1999; 99US-0134768P.
XX PR
19-MAY-1999; 99US-0134941P.
XX PR
20-MAY-1999; 99US-0135124P.
XX PR
21-MAY-1999; 99US-0135353P.
XX PR
24-MAY-1999; 99US-0135629P.
XX PR
25-MAY-1999; 99US-0136021P.
XX PR
27-MAY-1999; 99US-0136392P.
XX PR
28-MAY-1999; 99US-0136782P.
XX PR
03-JUN-1999; 99US-0137222P.
XX PR
03-JUN-1999; 99US-0137528P.
XX PR
04-JUN-1999; 99US-0137502P.
XX PR
07-JUN-1999; 99US-0137724P.
XX PR
08-JUN-1999; 99US-0138094P.
XX PR
10-JUN-1999; 99US-0138540P.
XX PR
10-JUN-1999; 99US-0138847P.
XX PR
14-JUN-1999; 99US-0139119P.
XX PR
16-JUN-1999; 99US-0139452P.
XX PR
16-JUN-1999; 99US-0139453P.
XX PR
17-JUN-1999; 99US-0139452P.
XX PR
18-JUN-1999; 99US-0139454P.
XX PR
18-JUN-1999; 99US-0139455P.
XX PR
18-JUN-1999; 99US-0139456P.
XX PR
18-JUN-1999; 99US-0139457P.
XX PR
18-JUN-1999; 99US-0139458P.
XX PR
18-JUN-1999; 99US-0139459P.
XX PR
18-JUN-1999; 99US-0139460P.
XX PR
18-JUN-1999; 99US-0139461P.
XX PR
18-JUN-1999; 99US-0139462P.
XX PR
18-JUN-1999; 99US-0139463P.
XX PR
18-JUN-1999; 99US-0139750P.
XX PR
18-JUN-1999; 99US-0139763P.
XX PR
21-JUN-1999; 99US-0139817P.
XX PR
22-JUN-1999; 99US-0139899P.
XX PR
23-JUN-1999; 99US-0140353P.
XX PR
23-JUN-1999; 99US-0140354P.
XX PR
24-JUN-1999; 99US-0140895P.
XX PR
28-JUN-1999; 99US-0140823P.
XX PR
29-JUN-1999; 99US-0140991P.
XX PR
30-JUN-1999; 99US-0141287P.
XX PR
01-JUL-1999; 99US-0141842P.
XX PR
01-JUL-1999; 99US-0142154P.
XX PR
02-JUL-1999; 99US-0142055P.
XX PR
06-JUL-1999; 99US-0142390P.
XX PR
08-JUL-1999; 99US-0142803P.
XX PR
09-JUL-1999; 99US-0142920P.
XX PR
12-JUL-1999; 99US-0142977P.
XX PR
13-JUL-1999; 99US-0143542P.
XX PR
14-JUL-1999; 99US-0143624P.
XX PR
15-JUL-1999; 99US-0144005P.
XX PR
16-JUL-1999; 99US-0144085P.
XX PR
16-JUL-1999; 99US-0144086P.
XX PR
19-JUL-1999; 99US-0144325P.
XX PR
19-JUL-1999; 99US-0144331P.
XX PR
19-JUL-1999; 99US-0144332P.
XX PR
19-JUL-1999; 99US-0144333P.
XX PR
19-JUL-1999; 99US-0144334P.
```

PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144804P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145931P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147113P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147301P.  
PR 09-AUG-1999; 99US-0147433P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149923P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155133P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159323P.  
PR 14-OCT-1999; 99US-0159332P.

PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 22-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.7%; Score 48; DB 3; Length 311;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYFYEAANGTRDHHKG 19  
|||:|:|:|  
Db 119 SDEKLYKGIHYTDHKAG 136

## RESULT 38

AG34578  
ID AAG34578 standard; protein; 343 AA.

XX AC AAG34578;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 42097.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126264P.

PR 01-APR-1999; 99US-0126785P.

PR 06-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.



```
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.7%; Score 48; DB 3; Length 343;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFVEAANGTRDHKG 19
Db 151 SDEKLYKGHGTYDHRKAG 168

RESULT 39
AAG20945
ID AAG20945 standard; protein; 361 AA.
AC AAG20945;
XX AC
XX XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23321.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126284P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.

99US-0134219P.
99US-0134221P.
99US-0134370P.
99US-0134768P.
99US-0134941P.
99US-0135124P.
99US-0135353P.
99US-0135629P.
99US-0136021P.
99US-0136392P.
99US-0136782P.
99US-0137222P.
99US-0137528P.
99US-0137502P.
99US-0137724P.
99US-0138094P.
99US-0138540P.
99US-0138847P.
99US-0139119P.
99US-0139452P.
99US-0139453P.
99US-0139492P.
99US-0139454P.
99US-0139455P.
99US-0139456P.
99US-0139457P.
99US-0139458P.
99US-0139459P.
99US-0139460P.
99US-0139461P.
99US-0139462P.
99US-0139463P.
99US-0139750P.
99US-0139763P.
99US-0139817P.
99US-0139899P.
99US-0140353P.
99US-0140354P.
99US-0140695P.
99US-0140823P.
99US-0140991P.
99US-0141287P.
99US-0141842P.
99US-0142154P.
99US-0142055P.
99US-0142390P.
99US-0142803P.
99US-0142920P.
99US-0142977P.
99US-0143542P.
99US-0143624P.
99US-0144005P.
99US-0144085P.
99US-0144086P.
99US-0144325P.
99US-0144331P.
99US-0144332P.
99US-0144333P.
99US-0144334P.
99US-0144352P.
99US-0144632P.
99US-0144684P.
99US-0144814P.
99US-0145086P.
99US-0145088P.
99US-0145085P.
99US-0145087P.
99US-0145089P.
99US-0145192P.
99US-0145145P.
99US-0145218P.
99US-0145224P.
```





PR 24-MAY-1999; 99US-0135623P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138549P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142902P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 26-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 23-OCT-1999; 99US-0161404P.  
PR 23-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 25-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.

PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.7%; Score 48; DB 3; Length 361;  
Best Local Similarity 50.0%; Fred. No. 12;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFYEAAANGTRDHKKG 19  
||| | : |||||  
Db 169 SDEKLYGIGHGYTDHKAG 186

Search completed: November 24, 2004, 09:21:39  
Job time : 84.8152 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:04 ; Search time 87.3587 Seconds  
(without alignments)  
125.141 Million cell updates/sec

Title: US-09-719-379A-1

Perfect score: 105

Sequence: 1 RSDYKFEANGTRDHKKG 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Query Length | ID           | Description         |
|------------|-------|-------|--------------|--------------|---------------------|
| 1          | 99    | 94.3  | 353          | 1 OM52_HAEIN | P38368 haemophilus  |
| 2          | 99    | 94.3  | 359          | 1 OM53_HAEIN | P45996 haemophilus  |
| 3          | 53    | 50.5  | 1211         | 2 Q76YV0     | Q76YV0 bacterioph   |
| 4          | 53    | 50.5  | 1211         | 2 AAQ17796   | AAQ17796 bacterioph |
| 5          | 52.5  | 50.0  | 360          | 2 O86254     | O86254 haemophilus  |
| 6          | 50    | 47.6  | 536          | 2 Q72BV4     | Q72BV4 desulfovibr  |
| 7          | 50    | 47.6  | 536          | 2 AAS96008   | AAS96008 desulfovi  |
| 8          | 48    | 45.7  | 304          | 2 Q9LNI9     | Q9LNI9 arabidopsis  |
| 9          | 48    | 45.7  | 343          | 2 Q8GX84     | Q8GX84 arabidopsis  |
| 10         | 48    | 45.7  | 378          | 2 Q8L789     | Q8L789 arabidopsis  |
| 11         | 48    | 45.7  | 378          | 2 Q9FNG6     | Q9FNG6 arabidopsis  |
| 12         | 48    | 45.7  | 678          | 2 Q7PZH8     | Q7PZH8 anopheles g  |
| 13         | 47.5  | 45.2  | 698          | 2 Q6HA27     | Q6HA27 trypanosoma  |
| 14         | 47    | 44.8  | 445          | 2 Q7YR14     | Q7YR14 thimolophus  |
| 15         | 46.5  | 44.3  | 345          | 1 YG3V YEAST | P53292 saccharomyc  |
| 16         | 46    | 43.8  | 339          | 2 Q9S579     | Q9S579 guillardia   |
| 17         | 46    | 43.8  | 478          | 2 Q88325     | Q88325 pseudomonas  |
| 18         | 46    | 43.8  | 593          | 2 Q7SCQ3     | Q7SCQ3 neurospora   |
| 19         | 46    | 43.8  | 670          | 2 Q9W474     | Q9W474 drosophila   |
| 20         | 46    | 43.8  | 853          | 2 Q8ALU3     | Q8ALU3 bacteroides  |
| 21         | 46    | 43.8  | 1116         | 1 YN94 YEAST | P53751 saccharomyc  |
| 22         | 46    | 43.8  | 1549         | 1 Y1R3 YEAST | P40438 saccharomyc  |
| 23         | 46    | 43.8  | 1549         | 1 YJW2 YEAST | P40890 saccharomyc  |
| 24         | 45.5  | 43.3  | 381          | 2 Q8KAK9     | Q8KAK9 chlorobium   |
| 25         | 45    | 42.9  | 274          | 2 Q7YIHO     | Q7YIHO oryza sativ  |
| 26         | 45    | 42.9  | 274          | 2 AAR87289   | AAR87289 oryza sat  |
| 27         | 45    | 42.9  | 443          | 2 Q7YQN5     | Q7YQN5 megaderma 1  |
| 28         | 45    | 42.9  | 449          | 2 Q87T32     | Q87T32 vibrio para  |
| 29         | 45    | 42.9  | 631          | 2 Q8THX4     | Q8THX4 methanosarc  |
| 30         | 45    | 42.9  | 996          | 2 Q6XPR4     | Q6XPR4 emericella   |
| 31         | 45    | 42.9  | 996          | 2 AAR02857   | AAR02857 emericell  |

|    |      |      |      |            |                    |
|----|------|------|------|------------|--------------------|
| 32 | 45   | 42.9 | 2633 | 2 Q7R5E0   | Q7R5E0 giardia lam |
| 33 | 45   | 42.9 | 2734 | 2 Q86BA3   | Q86BA3 drosophila  |
| 34 | 45   | 42.9 | 2734 | 2 Q8MDK0   | Q8MDK0 drosophila  |
| 35 | 45   | 42.9 | 3002 | 2 Q86BA2   | Q86BA2 drosophila  |
| 36 | 45   | 42.9 | 3002 | 2 Q8MUJ9   | Q8MUJ9 drosophila  |
| 37 | 45   | 42.9 | 4723 | 2 Q86BA1   | Q86BA1 drosophila  |
| 38 | 45   | 42.9 | 4723 | 2 Q8MUJ8   | Q8MUJ8 drosophila  |
| 39 | 44.5 | 42.4 | 364  | 2 Q6PSK1   | Q6PSK1 pasteurella |
| 40 | 44.5 | 42.4 | 364  | 2 Q6PSK5   | Q6PSK5 pasteurella |
| 41 | 44.5 | 42.4 | 364  | 2 AAS91717 | AAS91717 pasteur   |
| 42 | 44.5 | 42.4 | 364  | 2 AAS91721 | AAS91721 pasteur   |
| 43 | 44.5 | 42.4 | 364  | 2 AAS91725 | AAS91725 pasteur   |
| 44 | 44.5 | 42.4 | 613  | 2 Q7REA0   | Q7REA0 plasmodium  |
| 45 | 44   | 41.9 | 308  | 2 O28835   | O28835 archaeglob  |

#### ALIGNMENTS

##### RESULT 1

OM52\_HAEIN ID OM52\_HAEIN STANDARD; PRT; 353 AA.  
AC P38368;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Outer membrane protein P5 precursor (OMP P5).  
GN Name=ompA; Synonyms=ompP5;  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-34.  
RC STRAIN=1613 / Serotype B;  
RX MEDLINE=93366472; PubMed=8359929;  
RA Munson R.S. Jr., Grass S., West R.;  
RT "Molecular cloning and sequence of the gene for outer membrane protein P5 of Haemophilus influenzae."  
RL Infect. Immun. 61:4017-4020(1993).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
CC -1- SIMILARITY: Belongs to the ompA family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; L20309; AAA03346.1; -.  
HSP; P02934; 1QJP.  
DR InterPro; IPR006664; Bac OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR006690; OMPA LINK.  
DR InterPro; IPR000498; OmpA\_tmem.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA membrane; 1.  
DR PRINTS; PR01021; OMPADOMAIN.  
DR PRODOM; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OMPA; 1.  
KW Direct protein sequencing; Outer membrane; Porin; Signal;  
Transmembrane.  
FT SIGNAL 1 21  
FT CHAIN 22 353 Outer membrane protein P5.  
FT DISULFID 326 338 By similarity.  
FT DOMAIN 272 316 OmpA-like.  
SQ SEQUENCE 353 AA; 37594 MW; E58A659E786D0DF7 CRC64;  
Query Match 94.3%; Score 99; DB 1; Length 353;  
Best Local Similarity 94.7%; Pred. No. 3.2e-08;

```
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHKKG 19
|||||
Db 132 RSDYKFYEDANGTRDHKKG 150

RESULT 2
OM53 HABIN
ID_OM53_HABIN STANDARD; PRT; 359 AA.
AC P45996;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Outer membrane protein P5 precursor (OMP P5) (Fimbrin).
GN Name=ompA; Synonyms=ompP5;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NT1128; PubMed=7909539;
RA MEDLINE=9422575;
RA Sirakova T., Kolattukudy P.E., Murwin D., Billy J., Leake E., Lim D.,
RA Denaria T., Bakaletz L.;
RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
RT pathogenesis of and protection against otitis media and relatedness of
RT the fimbrin subunit to outer membrane protein A.";
RL Infect. Immun. 62:2002-2020(1994).
CC -!- FUNCTION: Acts as a fimbriae subunit.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: Belongs to the ompA family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L08448; AAA24959.1; -.
DR HSSP; P02934; 1BXW.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR002368; OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OMPA_LIKE.
DR InterPro; IPR000498; OmpA_tmcm.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01389; OmpA membrane; 1.
DR PRODOM; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OMPA; FALSE_NEG.
DR Direct protein sequencing; Fimbrin; Outer membrane; Porin; Signal;
KW Transmembrane.
FT SIGNAL 1 21
FT CHAIN 22 359 Outer membrane protein P5.
FT DISULFID 332 344 By similarity.
FT DOMAIN 278 322 OmpA-like.
SQ SEQUENCE 359 AA; 38340 MW; 576B1C59B4818C37 CRC64;

Query Match 94.3%; Score 99; DB 1; Length 359;
Best Local Similarity 94.7%; Pred. No. 3.2e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHKKG 19
|||||
Db 138 RSDYKFYEDANGTRDHKKG 156

RESULT 3
Q76YV0

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHKKG 19
|||||
Db 132 RSDYKFYEDANGTRDHKKG 150

ID Q76YV0 PRELIMINARY; PRT; 1211 AA.
AC Q76YV0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AehIorf13ic;
OS Bacteriophage AehI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=227470;
RN [1]
RP SEQUENCE FROM N.A.
RA Petrov V., Nolan J., Bertrand C., Letarov A.V., Kirsch H.M.,
RA Karam J.D.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Petrov V., Nolan J., Karam J.D.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY266303; AAQ17796.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1211 AA; 137634 MW; 6447E705BB2A9A88 CRC64;

Query Match 50.5%; Score 53; DB 2; Length 1211;
Best Local Similarity 43.5%; Pred. No. 9.2;
Matches 10; Conservative 5; Mismatches 4; Indels 4; Gaps 1;

Qy 1 RSDYKFYEAAANG-----TRDHKKG 19
:::|::|
Db 773 KNGKFFNAENGKSISVRDHKQG 795

RESULT 4
AAQ17796 PRELIMINARY; PRT; 1211 AA.
ID AAQ17796
AC AAQ17796;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN AehIorf13ic.
OS Bacteriophage AehI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=227470;
RN [1]
RP SEQUENCE FROM N.A.
RA Petrov V., Nolan J., Bertrand C., Letarov A.V., Kirsch H.M.,
RA Karam J.D.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Petrov V., Nolan J., Karam J.D.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY266303; AAQ17796.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1211 AA; 137634 MW; 6447E705BB2A9A88 CRC64;

Query Match 50.5%; Score 53; DB 2; Length 1211;
Best Local Similarity 43.5%; Pred. No. 9.2;
Matches 10; Conservative 5; Mismatches 4; Indels 4; Gaps 1;

Qy 1 RSDYKFYEAAANG-----TRDHKKG 19
:::|::|
Db 773 KNGKFFNAENGKSISVRDHKQG 795

RESULT 5
O86254 PRELIMINARY; PRT; 360 AA.
ID O86254
AC O86254;
```

DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Outer membrane protein (Fragment).  
GN Name=omp;  
OS Haemophilus sp.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=740;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16N;  
RX MEDLINE=99081716; PubMed=9864189;  
RA Goussset N., Rosenau A., Sizaret P.Y., Quentin R.;  
RT "Nucleotide sequences of genes coding for fimbrial proteins in a  
RT cryptic genospecies of Haemophilus spp. isolated from neonatal and  
RT genital tract infections.";  
RL Infect. Immun. 67:8-15(1999).  
CC -1- SIMILARITY: Belongs to the ompA family.  
DR EMBL; AJ007317; CAA07454.1; -.  
DR GO; GO:0016021; C: integral to membrane; IEA.  
DR GO; GO:000279; C: outer membrane (sensu Gram-negative Bacteria); IEA.  
DR GO; GO:0005198; F: structural molecule activity; IEA.  
DR InterPro; IPR006664; Bac\_OmpA.  
DR InterPro; IPR002368; OmpA.  
DR InterPro; IPR006665; OmpA/MotB.  
DR InterPro; IPR006690; OmpA\_LIKE.  
DR InterPro; IPR000498; OmpA\_tmtn.  
DR Pfam; PF00691; OmpA; 1.  
DR Pfam; PF01389; OmpA membrane; 1.  
DR PRINTS; PR01021; OMPADOMAIN.  
DR PRINTS; PR01022; OUTRMMERANE.  
DR PRODOM; PD000930; OmpA/MotB; 1.  
DR PROSITE; PS01068; OmpA; 1.  
FT NON TER 360 360  
SQ SEQUENCE 360 AA; 38415 MW; A3209155051CDD69 CRC64;  
  
Query Match 50.0%; Score 52.5; DB 2; Length 360;  
Best Local Similarity 78.6%; Pred. No. 2.9;  
Matches 11; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 RSDYKFEVANGTR 14  
DB 140 RSDYKFD-ANGAR 152  
  
RESULT 6  
QY 72BV4 PRELIMINARY; PRT; 536 AA.  
AC Q72BV4  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Metallo-beta-lactamase family protein.  
GN OrderedLocusNames=DVU1530;  
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;  
OC Desulfovibrionaceae; Desulfovibrio.  
OX NCBI\_TaxID=882;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC PubMed=15077118; DOI=10.1038/nbt959;  
RA Kolonay J.F., Eiseen J.A., Ward N., Methe B.A., Brinkac L.M.,  
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,  
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,  
RA Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,  
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,  
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;  
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium  
RT Desulfovibrio vulgaris Hildenborough.";  
RL Nat. Biotechnol. 22:554-559(2004).  
DR EMBL; AE017314; AAS96008.1; -.  
DR TIGR; DVU1530; -.  
SQ SEQUENCE 536 AA; 59893 MW; C6D54A537BBCC286 CRC64;  
  
Query Match 47.6%; Score 50; DB 2; Length 536;  
Best Local Similarity 56.2%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 SDYKFEVANGTRDHK 17  
DB 206 ADYLFLESTYGDGRDHK 221  
  
RESULT 8  
QY 99LN19 PRELIMINARY; PRT; 304 AA.  
AC Q99LN19  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative zinc finger protein.  
GN Name=P6F3.14;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]

DR EMBL; AE017314; AAS96008.1; -.  
DR TIGR; DVU1530; -.  
DR InterPro; IPR001279; Blactmase-like.  
DR InterPro; IPR011108; RMBL.  
DR Pfam; PF00753; Lactamase\_B; 1.  
DR Pfam; PF07521; RMBL; 1.  
KW Complete proteome.  
SQ SEQUENCE 536 AA; 59893 MW; C6D54A537BBCC286 CRC64;  
  
Query Match 47.6%; Score 50; DB 2; Length 536;  
Best Local Similarity 56.2%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 SDYKFEVANGTRDHK 17  
DB 206 ADYLFLESTYGDGRDHK 221  
  
RESULT 7  
QY 99LN19 PRELIMINARY; PRT; 536 AA.  
AC AAS96008  
DT 27-APR-2004 (TrEMBLrel. 27, Created)  
DT 27-APR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 11-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
DE Metallo-beta-lactamase family protein.  
GN DVU1530.  
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;  
OC Desulfovibrionaceae; Desulfovibrio.  
OX NCBI\_TaxID=882;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC PubMed=15077118;  
RA Heidelberg J.F., Eiseen J.A., Ward N., Methe B.A., Brinkac L.M.,  
RA Kolonay J.F., Dodson R.J., Durkin A.S., Madupu R.,  
RA Daugherty S.C., DeBoy R.T., Fouts D.E., Haft D.H., Selengut J.,  
RA Nelson W.C., Sullivan S.A., Fouts D.E., Zafar N., Zhou L., Radune D.,  
RA Peterson J.D., Daviden T.M., Zafar N., Khouri H.M., Gill J., Utterback T.R.,  
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;  
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium  
RT Desulfovibrio vulgaris Hildenborough.";  
RL Nat. Biotechnol. 22:554-559(2004).  
DR EMBL; AE017314; AAS96008.1; -.  
DR TIGR; DVU1530; -.  
SQ SEQUENCE 536 AA; 59893 MW; C6D54A537BBCC286 CRC64;  
  
Query Match 47.6%; Score 50; DB 2; Length 536;  
Best Local Similarity 56.2%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 SDYKFEVANGTRDHK 17  
DB 206 ADYLFLESTYGDGRDHK 221  
  
RESULT 8  
QY 99LN19 PRELIMINARY; PRT; 304 AA.  
AC Q99LN19  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative zinc finger protein.  
GN Name=P6F3.14;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]

SEQUENCE FROM N.A.

RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
RA Altati H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,  
RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
RA Lenz C., Liu A., Liu S., Mukharsy N., Pham P., Sakano H., Shinn P.,  
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: Contains 1 RING-type zinc finger.

CC ENBL; AC023628; AAF97335.1; -.  
DR PUR; G86143; G86143.  
DR HSP; P38398; LJM7.  
DR GO; GO:000151; Cribiquitin ligase complex; IEA.  
DR GO; GO:0003676; F-nucleic acid binding; IEA.  
DR GO; GO:0004842; F-ubiquitin-protein ligase activity; IEA.  
DR GO; GO:0008270; F-zinc ion binding; IEA.  
DR GO; GO:0016567; P-protein ubiquitination; IEA.  
DR InterPro; IPR000571; Znf-CCCH.  
DR InterPro; IPR001841; Znf\_Ring.  
DR Pfam; PF000097; zf-C3HC4; 1.  
DR Pfam; PF00642; zf-CCCH; 1.  
DR SMART; SM00184; RING; 1.  
DR SMART; SM00356; Znf\_C3H1; 1.  
DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
DR PROSITE; PS00089; ZF\_RING\_2; 1.  
DR Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 304 AA, 34215 MW, 7C10707FA1642951D27 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 304;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFYEAANGTRDHKKG 19  
||| : |||  
Db 151 SDEKLYKGIGHGYTDHKAG 168

| RESULT 9 | Q8GX84   | PRELIMINARY; | PRT; | 343 AA. |
|----------|--|--------------|------|---------|
| ID       | Q8GX84   |              |      |         |
| AC       | Q8GX84;  |              |      |         |
| DT       | 01-MAR-2003 (TrEMBLrel. 23, Created)   |              |      |         |
| DT       | 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  |              |      |         |
| DT       | 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  |              |      |         |
| DE       | Hypochemical protein Atlg01350/F6F3_27.  |              |      |         |
| GN       | Name=Atlg01350/F6F3_27;  |              |      |         |
| OS       | Arabidopsis thaliana (Mouse-ear cress).  |              |      |         |
| OC       | Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; |              |      |         |
| OC       | eurosid II; Brassicales; Brassicaceae; Arabidopsi.   |              |      |         |
| OX       | NCBI_TaxID=3702;   |              |      |         |
| NC       | [1]  |              |      |         |
| NR       | SEQUENCE FROM N.A.   |              |      |         |
| RP       | Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,   |              |      |         |
| RA       | Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,   |              |      |         |
| RA       | Hayashizaki Y., Shinozaki K.,  |              |      |         |
| RL       | Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  |              |      |         |
| CC       | -1- SIMILARITY: Contains 1 RING-type zinc finger.  |              |      |         |
| DR       | EMBL; AK118376; BAC42988.1; -.   |              |      |         |
| DR       | HSSP; P15919; 1RMD.  |              |      |         |
| DR       | GO; GO:0000151; Cubiquitin ligase complex; IEA.  |              |      |         |
| DR       | GO; GO:0003676; F-nucleic acid binding; IEA.   |              |      |         |
| DR       | GO; GO:0004842; F-ubiquitin-protein ligase activity; IEA.  |              |      |         |
| DR       | GO; GO:0008270; F-zinc ion binding; IEA.   |              |      |         |
| DR       | GO; GO:0016567; P-protein ubiquitination; IEA.   |              |      |         |
| DR       | InterPro; IPR000571; Znf CCHC.   |              |      |         |
| DR       | InterPro; IPR001841; Znf ring.   |              |      |         |
| DR       | Pfam; PF000097; zf-C3HC4_1.  |              |      |         |
| DR       | Pfam; PF00842; zf-CCHC; 1.   |              |      |         |
| DR       | SMART; SM00184; RING; 1.   |              |      |         |
| DR       | SMART; SM00356; Znf_C3H1; 1.   |              |      |         |
| DR       | PROSITE; PS00518; ZF_RING_1; UNKNOWN_1.  |              |      |         |
| DR       | PROSITE; PS50089; ZF_RING_2; 1.  |              |      |         |

KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 343 AA; 38719 MW; E333D70369C31A83 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 343;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 9; Conservative 2; Mismatches 7; Indels

Qy 2 SDYKFYEANGTRDHKKG 19  
||| : |||  
Db 151 SDEKLYKGIGHGYTTHKAG 16

RESULT 10  
Q8L789

|        |              |      |         |
|--------|--------------|------|---------|
| Q8L789 | PRELIMINARY; | PRT; | 378 AA. |
|--------|--------------|------|---------|

|    |                                |   |
|----|--------------------------------|---|
| AC | Q8L789;                        |   |
| DT | 01-OCT-2002                    | (TrEMBLrel. 22, Created)                |
| DT | 01-OCT-2002                    | (TrEMBLrel. 22, Last sequence update)   |
| DT | 05-JUL-2004                    | (TrEMBLrel. 27, Last annotation update) |
| DE | Hypothetical protein A5506420. |   |

OS *Arabis* *Arabis thaliana* (Mouse-ear cress).  
 OC *Eukaryota*; Viridiplantae; Streptophyta;  
 OC *Spermatophyta*; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC *eurosid* II; Brassicales; Brassicaceae; Arabidopses.  
 NCBI\_TaxID=3702;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP

REFERENCE FROM N.A.

RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., D  
RA Carncini P., Chen H., Cheuk R., Chan M.M., Chang C.H., D  
RA Deng J.M., Havaehizaki Y., Hsuan V.W., Lee J.M., Ishida  
RA Kawai J., Kim C.J., Narusaka M., Quach H.L., Sakurai T.,  
RA Seki M., Shinn P., Tang C.C., Toroumi M., Wallender E.K.,  
RA Wu H.C., Yamada K., Yu G., Yuan S., Shinozaki K., Becker  
RA Theologis A., Davis R.W.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]

RN [Z]  
RP SEQUENCE FROM N.A.

RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: Contains 1 RING-type zinc finger.

DR EMBL; AY136406; AAM97072.1; -.

DR EMBL; BT000229; AAN15548.1; -.

DR GO:0000151; C:ubiquitin ligase complex; IEA.

DR GO: 0003676: F:nucleic acid binding; IEA.

DR GO:0005070; E:ubiquitin-protein ligase activity: IEA;  
DR GO:0005070; E:ubiquitin-protein ligase activity: IEA;

DR GQ: GQ:0008270: F:zinc ion binding: IEA.

DR GQ:0008270: F:ZINC ION BINDING; TEA.  
DR GQ:0016567: P:protein ubiquitination; TEA

DR InterPro: IPR000571: ZnF CCH.

```

DK      INCELFIO; IFR000371; ZML_CCH:
DR      Interpro: IPR001841: Znf_ying

```

DK INCELFIO; IFR001841; ZML\_ILNG.  
DP Pfam: PF000097. zf-C3HCA.1

DR PFAM; PF00097; ZI-C3HC4; 1.  
DP pfam: PF00642; ZF-CCH; 1

DR PFAM; PF00642; ZI-CUCH; 1.  
DR CMAPT. CM00194. PING. 1

DR SMART; SM00184; RING; 1.  
DR SMART; SM00250; 2-5 CUBIC 1

DR SMART; SM00356; ZnF C3H1; 1.

```

DR      PROSITE; PS00518;
DR      ZF_RING_1; UniProt

```

DR PROSITE; PS50089; ZF\_RING\_2; 1.

KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.

SQ SEQUENCE 378 AA; 42518 MW; 142E4A6534BECA4D CRC64;

Query Match 45.7%; Score 48; DB 2; Length 378;

Best Local Similarity 50.0%; Pred. No. 18;

Qy 2 SDYKFYEANGTRDHKKG 19  
|||:|  
Db 186 SDEKLYKGIHGYTDHKAG 20

## RESULT 11

Q9FNG6 PRELIMINARY; PRT; 378 AA.  
 AC Q9FNG6;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Similarity to zinc finger protein (Hypothetical protein).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98069011; PubMed=9405937;  
 RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.  
 RT Sequence features of the regions of 1,044,062 bp covered by thirteen  
 RT physically assigned P1 clones.";  
 RL DNA Res. 4:291-300(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22088475; PubMed=12093376;  
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RT annotation.";  
 RL Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.  
 DR EMBL; AB006700; BAB08964.1; -.  
 DR EMBL; AY087435; AAM67329.1; -.  
 DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0016567; F:protein ubiquitination; IEA.  
 DR InterPro; IPR00571; Znf\_CCHC.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR Pfam; PF00642; zf-CCCH; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00356; Znf\_C3H1; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
 DR PROSITE; PS0089; ZF\_RING\_2; 1.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 378 AA; 42460 MW; 173D71BB88A3FE2D CRC64;

Query Match 45.7%; Score 48; DB 2; Length 378;  
 Best Local Similarity 50.0%; Pred. No. 18;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKFEYAANGTRDHKK 19

Db 186 SDEKLYGIRGYTDHKAG 203

## RESULT 12

Q7PZH8 PRELIMINARY; PRT; 678 AA.  
 AC Q7PZH8;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE AGCP9402 (Fragment).  
 GN Name=agCC54458; ORFNames=ENSGANGG00000012855;  
 OS Anopheles gambiae str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAB01008986; EAA00289.1; -.  
 DR InterPro; IPR009613; DUF1222.  
 DR Pfam; PF06762; DUF1222; 1.  
 DR NON\_TER 1  
 FT NON\_TER 678 678  
 SQ SEQUENCE 678 AA; 77111 MW; 97E9FDD34B130D02 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 678;  
 Best Local Similarity 50.0%; Pred. No. 34;  
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 RSDYKFEYAANGTRDHKK 18

Db 532 RPDFKFYDAAGSKSDAQK 549

## RESULT 13

Q6HA27 PRELIMINARY; PRT; 698 AA.  
 AC Q6HA27;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Prolyl oligopeptidase (EC 3.4.21.26).  
 GN Namespop;  
 OS Trypanosoma brucei.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5691;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Bastos I.M.D., Santana J.M., Grellier P.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ496456; CAD42967.1; -.  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR002470; Peptidase\_S9A.  
 DR InterPro; IPR004106; Peptidase\_S9A\_N.  
 DR InterPro; IPR002471; Pept\_S9\_AS.  
 DR InterPro; IPR000379; Ser\_estrs.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 DR Pfam; PF02897; Peptidase\_S9\_N; 1.  
 DR PRINTS; PR00862; PROLIGOPTASE.  
 DR PROSITE; PS00708; PRO\_ENDOPEP\_SER; UNKNOWN\_1.  
 KW Hydrolase.  
 SQ SEQUENCE 698 AA; 77597 MW; A969F75972E45910 CRC64;

Query Match 45.2%; Score 47.5; DB 2; Length 698;  
 Best Local Similarity 43.5%; Pred. No. 43;  
 Matches 10; Conservative 5; Mismatches 3; Indels 5; Gaps 1;

QY 2 SDYKFEYAANGTR----DHKK 19

Db 433 TEQFYNSADGTRIPWFIHRKG 455

## RESULT 14

Q7YR14 PRELIMINARY; PRT; 445 AA.  
 ID Q7YR14  
 AC Q7YR14;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Apolipoprotein B (Fragment).

```

OS Rhinolophus creaghi.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Rhinolophidae;
OC Rhinolophinae; Rhinolophus.
OX NCB1_TaxID=178895;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22761261; PubMed=12878460;
RA Amine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT "A new phylogenetic marker, apolipoprotein B, provides compelling
RT evidence for eutherian relationships.";
RL Mol. Phylogenet. Evol. 28:225-240(2003).
DR EMBL; AY243371; AAP50759.1; -.
KW Lipoprotein.
FT NON TER 1
FT NON TER 445
FT NON TER 445
SQ SEQUENCE 445 AA; 49395 MW; 6ECCAB23CB347E73 CRC64;

Query Match 44.8%; Score 47; DB 2; Length 445;
Best Local Similarity 69.2%; Pred. No. 32;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 KFYEAANGTRDHK 17
Db 371 KLYSAATGTVDHK 383

RESULT 15
YGV3_YEAST STANDARD; PRT; 345 AA.
ID YGV3_YEAST
AC P53292;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hypothetical 39.6 kDa protein in GTR2-KRE11 intergenic region.
GN OrderedLocNames=YGR165W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCB1_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RC MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; Z72950; CAA97189.1; -.
DR PIR; S64476; S64476.
DR GenOnline; 141477; -.
DR SGD; S0003397; YGR165W.
DR GO; GO:0005763; C:mitochondrial small ribosomal subunit; IPI.
DR GO; GO:0003735; F:structural constituent of ribosome; IPI.
DR GO; GO:0009060; P:aerobic respiration; IEP.
DR GO; GO:0006412; P:protein biosynthesis; IC.
KW Hypothetical protein.
SQ SEQUENCE 345 AA; 39575 MW; 4E363E30F5056329 CRC64;

Query Match 44.3%; Score 46.5; DB 1; Length 345;
Best Local Similarity 40.7%; Pred. No. 29;
Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Qy 1 RSDYKF-----YEAANGTRDHHK 18

```

```

Db 304 RSOYKFTNAKVGKVGKVGSGNRDNKK 330
|| ||| | : |||
304 RSOYKFTNAKVGKVGKVGSGNRDNKK 330

RESULT 16
Q98S79 PRELIMINARY; PRT; 339 AA.
AC Q98S79;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA-directed RNA polymerase 40k chain.
GN Name=rpas;
OS Guillardia theta (Cryptomonas phi).
OC Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCB1_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096(2001).
DR EMBL; AF083031; AAK39703.1; -.
DR PIR; C90126; C90126.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0046983; F:protein dimerization activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR011261; RNAP dimerization.
DR InterPro; IPR011262; RNAP insert.
DR InterPro; IPR009025; RNAP_RBP1-like.
DR InterPro; IPR011263; RNAP_RpoA_D_Rpb3.
DR InterPro; IPR001514; RNA_PolD.
DR Pfam; PF01000; RNA_Pol_A_bac; 1.
DR Pfam; PF01193; RNA_Pol_L; 1.
DR SMART; SM00662; RPOLD; 1.
DR PROSITE; PS00446; RNA_POL_D_30KD; 1.
KW DNA-directed RNA polymerase; Nucleomorph.
SQ SEQUENCE 339 AA; 39207 MW; 5CDD56AC58F1A3CA CRC64;

Query Match 43.8%; Score 46; DB 2; Length 339;
Best Local Similarity 43.8%; Pred. No. 35;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKPYEAANGTRDHHK 19
|| : |||
2 YKIFDTINNLKDNKK 17

RESULT 17
Q98325 PRELIMINARY; PRT; 478 AA.
AC Q98325;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 2-oxoglutarate dehydrogenase, E3 component, lipamide
DE dehydrogenase.
GN Name=lpdA; OrderedLocNames=PSPT02201;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCB1_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DC3000;
RC MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,

```



RA Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,  
RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,  
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
RA Bender C.L., White O., Fraser C.M., Collmer A.,  
RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
RT Pseudomonas syringae pv. tomato DC3000.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
CC -1- CATALYTIC ACTIVITY: Protein N(6)-(dihydrolipoyl)lysine + NAD(+) =  
CC protein N(6)-(lipoyl)lysine + NADH.  
CC -1- COFACTOR: Binds 1 FAD per subunit (By similarity).  
CC -1- COPACTOR: FAD (By similarity).  
CC -1- MISCELLANEOUS: The active site is a redox-active disulfide bond  
CC (By similarity).  
CC -1- SIMILARITY: Belongs to the class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
DR EMBL: AE016863; AA055717.1; -.  
DR HSP; P14218; ILPP.  
DR TIGR; PSTO2201; -.  
DR GO; GO:0005737; Cytoplasm; IEA.  
DR GO; GO:0004148; F:dihydrolipoyl dehydrogenase activity; IEA.  
DR GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006096; P:glycolysis; IEA.  
DR InterPro; IPR001327; FAD\_Pyr\_redox.  
DR InterPro; IPR000815; Hg\_reductase.  
DR InterPro; IPR006258; Lipamide dh.  
DR InterPro; IPR001100; Pyr\_redox\_dim.  
DR InterPro; IPR004099; Pyr\_redox\_dim.  
DR Pfam; PF00070; Pyr\_redox; 1.  
DR Pfam; PF02852; Pyr\_redox\_dim; 1.  
DR PRINTS; PR00368; FADPNR.  
DR PRINTS; PR00945; HGRDTASE.  
DR PRINTS; PR00411; PNDRTASSI.  
DR ProDom; PD000139; FAD\_Pyr\_redox; 1.  
DR TIGRFAMs; TIGR01350; Lipamide DH; 1.  
DR TrEMBL; F000076; PYRIDINE REDOX 1; 1.  
DR PROSITE; PS00076; PYRIDINE REDOX; NAD; Oxidoreductase;  
KW Complete proteome; FAD; Flavoprotein; NAD;  
KW Redox-active center.  
SQ SEQUENCE 478 AA; 50026 MW; 912C9934A31D0505 CRC64;

Query Match 43.8%; Score 46; DB 2; Length 478;  
Best Local Similarity 60.0%; Pred. No. 51;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDVKFYEAANGTRDH 16  
Db 63 SSWKIFYEAKNGFAVH 77

RESULT 18

ID Q7SCQ3 PRELIMINARY; PRT; 593 AA.

AC Q7SCQ3;

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Predicted protein.

GN Name=NCU03250.1;

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OR74A;

RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,

RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,

RA Elkins T., Engle R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,

RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M.,

RA Seitzrennikoff C.P., Kinsey J.A., Braun E.L., Zeiter A., Schulte U.,

RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,  
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,  
RA Kamal M., Kamysellis M., Mauceli E., Bielek C., Rudd S., Frishman D.,  
RA Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,  
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmari S.A.,  
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
RA Varden O., Planann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
RA Natvig D.O., Alex L.A., Mannhaupt G., Eboole D.J., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,  
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";  
RL Nature 0:0-0(2003).  
CC -1- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AABX01000086; EAA34529.1; -.  
DR InterPro; IPR008266; Tyr\_kinase AS.  
DR PROSITE; PS00109; PROTEIN KINASE TVR; UNKNOWN 1.  
SQ SEQUENCE 593 AA; 67688 MW; DB0893E77A4F8B7F CRC64;

Query Match 43.8%; Score 46; DB 2; Length 593;  
Best Local Similarity 75.0%; Pred. No. 65;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 7 YEAAAGTRDHKK 18

Db 244 YDAANGTRAKKK 255

RESULT 19

ID Q9W474 PRELIMINARY; PRT; 670 AA.

AC Q9W474;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE CG4064-PA.

GN ORFNames=CG4064;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Merkulov G., Milshina N., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,



FT TRANSMEM 934 957 Potential.  
 FT CARBOHYD 335 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 336 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 553 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 846 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 985 N-linked (GlcNAc. . .) (Potential).  
 SQ SEQUENCE 1116 AA; 125199 MW; C0361878FE4DAB90 CRC64;

Query Match 43.8%; Score 46; DB 1; Length 1116;  
 Best Local Similarity 52.9%; Pred. No. 1.3e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DYKFEAANGTRDRHKKG 19  
 |||:|:|:|  
 Db 857 DYNFYKASDGTCKLVKG 873

RESULT 22  
 YIR3 YEAST  
 ID\_YIR3\_YEAST STANDARD; PRT; 1549 AA.  
 AC P40438;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Putative membrane glycoprotein YIL173W precursor.  
 GN OrderedLocusNames=YIL173W;  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 [1]

SEQUENCE FROM N.A.  
 STRAIN=S288C / AB972;  
 MEDLINE=97313266; PubMed=9169870;  
 RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,  
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,  
 RA Harris D.B., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,  
 RA Moulton S., Odell C., Pearson D., Rajadream M.A., Rice P., Rowley N.,  
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";  
 RL Nature 387:84-87(1997).  
 CC -1- SIMILARITY: BELONGS TO THE PEP1 FAMILY OF MEMBRANE GLYCOPROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z46921; CAA87019.1; -.  
 CC GERMOnline; 139708; -.  
 CC SCD; S0001435; VTH1.  
 CC GO; GO:0005768; C:Endosome; IDA.  
 CC GO; GO:0006896; P:Golgi to vacuole transport; IMP.  
 CC InterPro; IPR002860; Glyco\_hydro\_BNR.  
 CC InterPro; IPR011040; Sialidase.  
 CC InterPro; IPR006581; VPS10.  
 CC Pfam; PF02012; BNR; 12.  
 CC SMART; SM00602; VPS10; 2.  
 CC Glycoprotein; Hypothetical protein; Signal; Transmembrane.  
 CC SIGNAL 1 21  
 FT CHAIN 22 1549 Putative membrane glycoprotein YIL173W.  
 FT TRANSMEM 1370 1390 Potential.  
 FT CARBOHYD 479 479 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 769 769 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 986 986 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 1418 1418 N-linked (GlcNAc. . .) (Potential).  
 SQ SEQUENCE 1549 AA; 174426 MW; B2737747027B4E4F CRC64;

Query Match 43.8%; Score 46; DB 1; Length 1549;  
 Best Local Similarity 52.9%; Pred. No. 1.9e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DYKFEAANGTRDRHKKG 19  
 |||:|:|:|  
 Db 1290 DYNFYKASDGTCKLVKG 1306

RESULT 24  
 Q8KAK9  
 ID Q8KAK9 PRELIMINARY; PRT; 381 AA.  
 AC Q8KAK9  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)

Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DYKFEAANGTRDRHKKG 19  
 |||:|:|:|  
 Db 1290 DYNFYKASDGTCKLVKG 1306

RESULT 23  
 YJW2 YEAST  
 ID\_YJW2\_YEAST STANDARD; PRT; 1549 AA.  
 AC P40890;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Putative membrane glycoprotein YJL222W precursor.  
 GN OrderedLocusNames=YJL222W; ORFNames=J0213, HRC1549;  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 [1]

SEQUENCE FROM N.A.  
 STRAIN=S288C;  
 MEDLINE=95242842; PubMed=7725802;  
 RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,  
 RA Hilger F.;  
 RA "Sequence analysis of a 40.2 kb DNA fragment located near the left  
 RA telomere of yeast chromosome X.";  
 RL Yeast 10:1657-1662(1994).  
 CC -1- SIMILARITY: BELONGS TO THE PEP1 FAMILY OF MEMBRANE GLYCOPROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z34098; CAA83988.1; -.  
 CC PIR; S50705; S50705.  
 CC GERMOnline; 141834; -.  
 CC SGD; S0003758; VTH2.  
 CC InterPro; IPR002860; Glyco\_hydro\_BNR.  
 CC InterPro; IPR011040; Sialidase.  
 CC InterPro; IPR006581; VPS10.  
 CC Pfam; PF02012; BNR; 12.  
 CC SMART; SM00602; VPS10; 2.  
 CC Glycoprotein; Hypothetical protein; Signal; Transmembrane.  
 CC SIGNAL 1 21  
 FT CHAIN 22 1549 Putative membrane glycoprotein YJL222W.  
 FT TRANSMEM 1370 1390 Potential.  
 FT CARBOHYD 479 479 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 769 769 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 986 986 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 1418 1418 N-linked (GlcNAc. . .) (Potential).  
 SQ SEQUENCE 1549 AA; 174400 MW; B360CF7927B4E4F CRC64;

Query Match 43.8%; Score 46; DB 1; Length 1549;  
 Best Local Similarity 52.9%; Pred. No. 1.9e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DYKFEAANGTRDRHKKG 19  
 |||:|:|:|  
 Db 1290 DYNFYKASDGTCKLVKG 1306

RESULT 24  
 Q8KAK9  
 ID Q8KAK9 PRELIMINARY; PRT; 381 AA.  
 AC Q8KAK9  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DT Hypothetical protein Cr2147.  
 GN OrderedLocusNames=Cr2147;  
 OS Chlorobium tepidum.  
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 OC Chlorobaculum.  
 OX NCBI\_TaxID=1097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TLS / ATCC 49652 / DSM 12025;  
 RX MEDLINE=22103685; PubMed=12093901; DOI=10.1073/pnas.132181499;  
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
 RA Dodson R.J., DeBoy R.T., Gwinn M.L., Nelson W.C., Haft D.H.,  
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.F., Yang F.,  
 RA Holt I.E., Umayam L.A., Mason T.M., Brenner M., Shea T.P.,  
 RA Parksey D.S., Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B.,  
 RA Radune D., Vamathevan J.J., Knouri H.M., White O., Gruber T.M.,  
 RA Ketchum K.A., Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
 RT "The complete genome sequence of Chlorobium tepidum TLS, a  
 RT photosynthetic, anaerobic, green-sulfur bacterium";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
 DR EMBL; AB012963; AAM73363.1; -.  
 DR TIGR; Cr2147; -.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 381 AA; 42377 MW; 2CDDDF72C48C479 CRC64;  
  
 Query Match 43.3%; Score 45.5; DB 2; Length 381;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 10; Conservative 1; Mismatches 2; Indels 3; Gaps 1;  
  
 Qy 1 RSDYKFVEAANGTRDH 16  
 ||: ||| |||||  
 Db 135 RSEVKLYE--GTRDH 147  
  
 RESULT 25  
 Q7Y1H0 PRELIMINARY; PRT; 274 AA.  
 AC Q7Y1H0;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein OSJNBA0057G07.1 (Hypothetical protein  
 DE OJ1365 D05.20).  
 GN Name=OSJNBA0057G07.1; Synonym=OJ1365 D05.20;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Buell R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC117988; AAF44689.1; -.  
 DR EMBL; AC096855; AAR87289.1; -.  
 DR Gramene; Q7Y1H0; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 274 AA; 30566 MW; 5AA6364076740387 CRC64;  
  
 Query Match 42.9%; Score 45; DB 2; Length 274;  
 Best Local Similarity 61.5%; Pred. No. 40;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
 Qy 2 SDYKFVEAANGTR 14  
 | | | | | | | |  
 Db 90 STYSFHDANGNR 102  
  
 RESULT 26  
 AAR87289 PRELIMINARY; PRT; 274 AA.  
 ID AAR87289  
 AC AAR87289;  
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein OJ1365\_D05.20.  
 GN OJ1365\_D05.20  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza; Oryza sativa.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC096855; AAR87289.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 274 AA; 30566 MW; 5AA6364076740387 CRC64;  
  
 Query Match 42.9%; Score 45; DB 2; Length 274;  
 Best Local Similarity 61.5%; Pred. No. 40;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
 Qy 2 SDYKFVEAANGTR 14  
 | | | | | | | |  
 Db 90 STYSFHDANGNR 102  
  
 RESULT 27  
 Q7YQN5 PRELIMINARY; PRT; 443 AA.  
 ID Q7YQN5  
 AC Q7YQN5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Apolipoprotein B 100 (Fragment).  
 GN Name=apoB-100;  
 OS Megaderma lyra (Indian false vampire).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Megadermatidae;  
 OC Megaderma.  
 OX NCBI\_TaxID=9413;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22761261; PubMed=12878460;  
 RA Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;  
 RT "A new phylogenetic marker, apolipoprotein B, provides compelling  
 RT evidence for eutherian relationships.";  
 RL Mol. Phylogenet. Evol. 28:225-240(2003).  
 KW EMBL; AF548428; AAF97384.1; -;  
 KM Lipoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 443 443  
 SQ SEQUENCE 443 AA; 49385 MW; 48774408BFA69DFF CRC64;  
 Query Match 42.9%; Score 45; DB 2; Length 443;  
 Best Local Similarity 69.2%; Pred. No. 69;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 5 KFYEAANGTRDHK 17  
 Db 369 KLYSAAAGTVDDK 381  
 RESULT 28  
 Q87T32  
 ID Q87T32 PRELIMINARY; PRT; 449 AA.  
 AC Q87T32;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DE Hypothetical protein VP0238.  
 GN OrderedLocusNames=VP0238;  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yaenagata T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 RT distinct from that of V. cholerae.";  
 RL Lancet 361:743-749(2003).  
 DR EMBL; AF005073; BAC58501.1; -;  
 DR InterPro; IPR001279; Blactmase-like.  
 DR InterPro; IPR011108; RMBBL.  
 DR Pfam; PF00753; Lactamase\_B; 1.  
 DR Pfam; PF07521; RMBBL; 1.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 449 AA; 50016 MW; 323EC3B4732F5AE2 CRC64;  
 Query Match 42.9%; Score 45; DB 2; Length 449;  
 Best Local Similarity 41.2%; Pred. No. 70;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 RSDYKFYEANGTRDHK 17  
 Db 195 RADYLFYESTYGNKEHE 211  
 RESULT 29  
 Q8THX4  
 ID Q8THX4 PRELIMINARY; PRT; 631 AA.  
 AC Q8THX4;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE Hypothetical protein MA4384.

GN OrderedLocusNames=MA4384;  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,  
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atcoor D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.S., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity.";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AE011158; AAM07726.1; -;  
 DR GO; GO:0050897; Fcobalt ion binding; IEA.  
 DR GO; GO:0008705; Fmethionine synthase activity; IEA.  
 DR GO; GO:0004853; Furophosphorylase decarboxylase activity; IEA.  
 DR GO; GO:0009086; Pmethionine biosynthesis; IEA.  
 DR GO; GO:0006779; Pporphyrin biosynthesis; IEA.  
 DR InterPro; IPR006158; B12-binding.  
 DR InterPro; IPR011007; B12-binding.  
 DR InterPro; IPR003759; CoMet synth B12.  
 DR InterPro; IPR010988; M synth B12 like.  
 DR InterPro; IPR000257; Uro decarboxyls.  
 DR Pfam; PF02310; B12-binding; 1.  
 DR Pfam; PF02607; B12-binding; 2; 1.  
 DR ProDom; PD003225; Uro decarboxyls; 1.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 631 AA; 68830 MW; AE0E5054F3FDD0CB CRC64;  
 Query Match 42.9%; Score 45; DB 2; Length 631;  
 Best Local Similarity 56.2%; Pred. No. 1e+02;  
 Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;  
 Qy 4 YKPY--EAANGTRDHK 17  
 Db 523 YTFVPDEAGKSKDKH 538  
 RESULT 30  
 Q6XPR4  
 ID Q6XPR4 PRELIMINARY; PRT; 996 AA.  
 AC Q6XPR4;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE BroA.  
 GN Name=broA;  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=bial niia4;  
 RX MEDLINE=92017851; PubMed=1922072;  
 RA Dowzer C.E., Kelly J.M.;  
 RT "Analysis of the creA gene, a regulator of carbon catabolite  
 RT repression in Aspergillus nidulans.";  
 RL Mol. Cell. Biol. 11:5701-5709(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=bial niia4;





RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Buseam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam K.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Liang D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT *melanogaster* euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celinker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: Contains 1 LIM zinc-binding domain.  
DR EMBL; AE003685; AAC01532.1; -.  
DR HSSP; P04006; LIML.  
DR FlyBase; FBgn0053208; MICAL.  
DR GO; GO:0007411; P:axon guidance; IMP.  
DR InterPro; IPR001715; Calponin-like.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR011051; RmlC\_like\_cupin.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00412; LIM; 1.  
DR ProDom; PD000094; LIM; 1.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00132; LIM; 1.  
DR PROSITE; PS00021; CH; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
DR PROSITE; PS00023; LIM\_DOMAIN\_2; 1.  
DR PROSITE; PS00023; LIM\_DOMAIN\_2; 1.  
KW LIM domain; Metal-binding; Zinc.  
SQ SEQUENCE 3002 AA; 332499 MW; 8C754C12F57E0337 CRC64;  
Query Match 42.9%; Score 45; DB 2; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 5.9e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;  
Qy 1 RSDYKFYEAANGTRDHKK 18  
Db 2805 RSDDESYY--ANETREHKK 2820  
||| : | ||| : |||  
RESULT 36  
Q8MUJ9  
ID Q8MUJ9 PRELIMINARY; PRT; 3002 AA.  
AC Q8MUJ9;  
DT 01-OCT-2002 (T-REMBLrel. 22, Created)  
DT 01-OCT-2002 (T-REMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (T-REMBLrel. 26, Last annotation update)  
DE MICAL medium isoform.  
GN Name=MICAL;  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Spheroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RP SEQUENCE FROM N.A.  
RA Terman J.R., Mao T., Pasterkamp R.J., Yu H.-H., Kolodkin A.L.;  
RT "MICALS, a family of conserved flavoprotein oxidoreductases, function  
RT in plexin-mediated axonal repulsion.";  
RL Cell 0:0-0(2002).  
CC -1- SIMILARITY: Contains 1 LIM zinc-binding domain.  
DR EMBL; AF520714; AAM55243.1; -.  
DR HSSP; P04006; LIML.  
DR FlyBase; FBgn0053208; MICAL.  
DR GO; GO:0007411; P:axon guidance; IMP.  
DR InterPro; IPR001715; Calponin-like.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR011051; RmlC\_like\_cupin.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00412; LIM; 1.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00132; LIM; 1.  
DR PROSITE; PS00021; CH; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
DR PROSITE; PS00023; LIM\_DOMAIN\_2; 1.  
KW LIM domain; Metal-binding; Zinc.  
SQ SEQUENCE 3002 AA; 332498 MW; FF3B9B44AF599346 CRC64;  
Query Match 42.9%; Score 45; DB 2; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 5.9e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;  
Qy 1 RSDYKFYEAANGTRDHKK 18



```

|||| : -||| :||| :|||
2805 RSSDESY--ANETREHK 2820

Db
RESULT 37
Q86BA1 ID Q86BA1 PRELIMINARY; PRT; 4723 AA.
AC Q86BA1
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG33208-PC (CG33208-pd)
GN Name=MICAL; ORFName=CG33208;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Esler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Klamel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheier F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S.E., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Taor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RL melanogaster euchromatic genome sequence."
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
RP SEQUENCE FROM N.A.

RX Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
DR EMBL; AE003685; AAC41533.1; -.
DR HSPB; P04006; IIML.
DR FLYBASE; FBgn0053208; MICAL.
DR GO; GO:0007411; P:axon guidance; IMP.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001781; LIM_
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF00307; CH; 1.
DR Pfam; PF00412; LIM; 1.
DR PRODOM; PD000094; LIM; 1.
DR SMART; SM00033; CH; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00021; CH; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00023; LIM DOMAIN 2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 4723 AA; 525042 MW; C96ECC28393C7E9F CRC64;

Query Match 42.9%; Score 45; DB 2; Length 4723;
Best Local Similarity 61.1%; Pred. No. 9.9e+02;
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 RSDYKFEVAAANGTRDHKK 18
Db 4526 RSDDESY--ANETREHK 4541
|||| : -||| :||| :|||

RESULT 38
Q8MUJ8 ID Q8MUJ8 PRELIMINARY; PRT; 4723 AA.
AC Q8MUJ8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MICAL long isoform.
GN Name=MICAL;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.

```

Search completed: November 24, 2004, 09:28:49  
Job time : 90.3587 secs

| Result No. | Score | Query |        |        | DB     | ID                             | Description |
|------------|-------|-------|--------|--------|--------|--------------------------------|-------------|
|            |       | Match | Length | Length |        |                                |             |
| 1          | 48    | 45.7  | 304    | 2      | G86143 | probable zinc finger           |             |
| 2          | 46.5  | 44.3  | 345    | 2      | S64476 | hypothetical protein           |             |
| 3          | 46    | 43.8  | 339    | 2      | C90126 | DNA-directed RNA polymerase    |             |
| 4          | 46    | 43.8  | 1116   | 2      | S63397 | probable membrane              |             |
| 5          | 46    | 43.8  | 1549   | 2      | S50705 | hypothetical protein           |             |
| 6          | 44    | 41.9  | 308    | 2      | D69429 | hypothetical protein           |             |
| 7          | 43    | 41.0  | 634    | 2      | T27959 | hypothetical protein           |             |
| 8          | 43    | 41.0  | 863    | 2      | T27958 | hypothetical protein           |             |
| 9          | 43    | 41.0  | 1579   | 2      | S25329 | carboxypeptidase Y             |             |
| 10         | 42.5  | 40.5  | 366    | 2      | E85729 | probable outer membrane        |             |
| 11         | 42.5  | 40.5  | 366    | 2      | D90888 | putative outer membrane        |             |
| 12         | 42    | 40.0  | 175    | 2      | G70157 | hypothetical protein           |             |
| 13         | 42    | 40.0  | 399    | 2      | S27879 | secretory protein              |             |
| 14         | 42    | 40.0  | 714    | 2      | F81962 | probable iron-regulated        |             |
| 15         | 42    | 40.0  | 809    | 2      | A12747 | conserved hypothetical protein |             |
| 16         | 42    | 40.0  | 857    | 1      | S31429 | S-receptor kinase              |             |
| 17         | 42    | 40.0  | 873    | 2      | H97528 | hypothetical protein           |             |
| 18         | 42    | 40.0  | 1353   | 2      | T00249 | ichi protein - ink             |             |
| 19         | 41    | 39.0  | 109    | 2      | T23176 | hypothetical protein           |             |
| 20         | 41    | 39.0  | 129    | 2      | T15083 | hypothetical protein           |             |
| 21         | 41    | 39.0  | 142    | 2      | G84201 | diadenosine tetrap             |             |
| 22         | 41    | 39.0  | 436    | 2      | T14532 | S-locus-specific g             |             |
| 23         | 41    | 39.0  | 429    | 2      | T14535 | S-locus-specific g             |             |
| 24         | 41    | 39.0  | 455    | 2      | F82345 | conserved hypothetical         |             |
| 25         | 41    | 39.0  | 474    | 2      | A81732 | glycogen synthase              |             |
| 26         | 41    | 39.0  | 519    | 2      | E90086 | t-complex protein              |             |
| 27         | 41    | 39.0  | 713    | 2      | A56268 | Fe-regulated protein           |             |
| 28         | 41    | 39.0  | 1015   | 2      | JC6552 | DNA topoisomerase              |             |
| 29         | 40.5  | 38.6  | 514    | 1      | SYBR8  | threonine synthase             |             |

A;Map position: 7R

Query Match 44.3%; Score 46.5; DB 2; Length 345;  
 Best Local Similarity 40.7%; Pred. No. 8.5;  
 Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;  
 QY 1 RSDYKPF-----YEAANGTRDHKK 18  
 ||||| | :|||:  
 Db 304 RSQYKFTNAKVGKVGRYGSGNRDNKK 330

RESULT 3  
 C90126  
 DNA-directed RNA polymerase 40k chain [imported] - Guillardia theta nucleomorph  
 C;Species: nucleomorph Guillardia theta  
 A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
 C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
 C;Accession: C90126  
 R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re  
 Nature 410, 1091-1096, 2001  
 A;Title: The highly reduced genome of an enslaved algal nucleus.  
 A;Reference number: A99082; MUID:11323671; PMID:11323671  
 A;Accession: C90126  
 A;Status: Preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-339 <DOU>  
 A;Cross-references: UNIPROT:Q98S79; GB:AF083031; NID:gi3794326; PIDN:AAK39703.1; GSPDB:G  
 C;Genetics:  
 A;Gene: xpa5  
 A;Map position: 3  
 A;Genome: nucleomorph  
 A;Superfamily: Saccharomyces cerevisiae DNA-directed RNA polymerase 40K chain  
 C;Keywords: nucleomorph

Query Match 43.8%; Score 46; DB 2; Length 339;  
 Best Local Similarity 43.8%; Pred. No. 10;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 YKFEAANGTRDHKKG 19  
 ||:|:|:|:  
 Db 2 YKIFDTNNLKDNKKG 17

RESULT 4  
 S63397  
 probable membrane protein YNR065c - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein N3539  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 09-Jul-2004  
 C;Accession: S63397  
 R;Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.  
 submitted to the Protein Sequence Database, April 1996  
 A;Reference number: S62944  
 A;Accession: S63397  
 A;Molecule type: DNA  
 A;Residues: 1-1116 <DUE>  
 A;Cross-references: UNIPROT:P53751; EMBL:Z71680; NID:gi302593; PID:e239847; PID:gi302594  
 A;Experimental source: strain S288C  
 C;Genetics:  
 A;Gene: MIPS:YNR065C  
 A;Cross-references: SGD:S0005348  
 A;Map position: 14R  
 C;Keywords: transmembrane protein  
 F;941-957/Domain: transmembrane #status predicted <TM1>  
 F;990-1006/Domain: transmembrane #status predicted <TM2>

Query Match 43.8%; Score 46; DB 2; Length 1116;  
 Best Local Similarity 52.9%; Pred. No. 32;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DYKFEAANGTRDHKKG 19  
 |||||:|:|:  
 Db 857 DYNFYKASDGTCKLVKG 873

RESULT 5  
 S50705  
 hypothetical protein YJL222w - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein HRC1549; hypothetical protein J0213; hypothetical  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 30-Sep-1991 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
 C;Accession: S50705; S57012; S50354; S45150  
 R;Vandenbol, M.; Durand, P.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.  
 Yeast 10, 1657-1662, 1994  
 A;Title: Sequence analysis of a 40.2 kb DNA fragment located near the left telomere of y  
 A;Reference number: S50701; MUID:95242842; PMID:7725802  
 A;Accession: S50705  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1549 <VAN>  
 A;Cross-references: UNIPROT:P40890; EMBL:Z34098; NID:g496934; PIDN:CAA83988.1; PID:g4969  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994  
 R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
 submitted to the Protein Sequence Database, September 1995  
 A;Reference number: S56835  
 A;Accession: S57012  
 A;Molecule type: DNA  
 A;Residues: 1-1549 <VAV>  
 A;Cross-references: EMBL:Z49497; NID:gl015614; PIDN:CAA89519.1; PID:gl015615; MIPS:YJL222:  
 R;Lye, G.; Bowman, S.; Churcher, C.  
 submitted to the EMBL Data Library, December 1994  
 A;Reference number: S50349  
 A;Accession: S50354  
 A;Molecule type: DNA  
 A;Residues: 1-1515, 'L', 1517-1549 <LYE>  
 A;Cross-references: GB:Z47047; EMBL:Z46921; NID:g603997; PID:g604002; GSPDB:GN00009; MIP:  
 C;Genetics:  
 A;Gene: SGD:VTH2; MIPS:YIL173w  
 A;Cross-references: SGD:S0003758  
 A;Map position: 10L; 9L  
 C;Keywords: transmembrane protein  
 F;2-18/Domain: transmembrane #status predicted <TM1>  
 F;1374-1390/Domain: transmembrane #status predicted <TM2>  
 F;1423-1439/Domain: transmembrane #status predicted <TM3>  
 Query Match 43.8%; Score 46; DB 2; Length 1549;  
 Best Local Similarity 52.9%; Pred. No. 45;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 DYKFEAANGTRDHKKG 19  
 ||||:|:|:  
 Db 1290 DYNFYKASDGTCKLVKG 1306

RESULT 6  
 D69429  
 hypothetical protein AFI437 - Archaeoglobus fulgidus  
 C;Species: Archaeoglobus fulgidus  
 C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
 C;Accession: D69429  
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Nature 390, 364-370, 1997  
 A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A;Reference number: A69250; MUID:98049343; PMID:9389475  
 A;Accession: D69429  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-308 <KLE>  
 A;Cross-references: UNIPROT:O28835; GB:AE001004; GB:AE000782; NID:g2689327; PIDN:AAB8981  
 C;Superfamily: [NiFe]-hydrogenase-3-type complex Eha, hydrophilic subunit Ehar  
 Query Match 41.9%; Score 44; DB 2; Length 308;







R;Dobson, R.  
submitted to the EMBL Data Library, June 1996  
A;Reference number: Z19703  
A;Accession: T231176  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-109 <WIL>  
A;Cross-references: UNIPROT:Q21089; EMBL:Z75543; PIDN:CAA99865.1; GSPDB:GNO0023; CESP:K0  
A;Experimental source: clone K01D12  
C;Genetics:  
A;Gene: CESP:K01D12.9  
A;Map position: 5  
A;Introns: 38/1  
C;Superfamily: hypothetical protein K01D12.8

Query Match 39.0%; Score 41; DB 2; Length 109;  
Best Local Similarity 53.3%; Pred. No. 21;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRD 15  
| | | | | : | | | | |  
Db 76 RKDTKCYEYEDGTD 90

RESULT 20  
T15083  
hypothetical protein E03D2.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T15083  
R;Beck, C.; Graves, T.; Becker, M.; Ozerky, P.  
submitted to the EMBL Data Library, December 1997  
A;Description: The sequence of C. elegans cosmid E03D2.  
A;Reference number: Z18290  
A;Accession: T15083  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-129 <BEC>  
A;Cross-references: UNIPROT:O44539; EMBL:AF039036; NID:g2736349; PID:g2736353; PIDN:AAB5  
A;Experimental source: strain Bristol N2; clone E03D2  
C;Genetics:  
A;Gene: CESP:E03D2.2  
A;Map position: 5  
A;Introns: 19/1; 39/2; 90/2; 110/2

Query Match 39.0%; Score 41; DB 2; Length 129;  
Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRD 15  
| | | | | : | | | | |  
Db 49 RAFYGFYNAGNSKED 63

RESULT 21  
G84201  
diadenosine tetraphosphate pyrophosphohydrolase [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: G84201  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: G84201  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-142 <STO>  
A;Cross-references: UNIPROT:Q9HS29; GB:AE004437; NID:g10580041; PIDN:AAG18979.1; GSPDB:G  
C;Genetics:

A;Gene: apa

Query Match 39.0%; Score 41; DB 2; Length 142;  
Best Local Similarity 47.1%; Pred. No. 27;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKFYEAAANGTRDHK 17  
| | | | | : | | | | |  
Db 69 RDEYDVYVFEANGDRHK 85

RESULT 22  
T14532  
S-locus-specific glycoprotein - wild cabbage (fragment)  
N;Alternate names: S glycoprotein  
C;Species: Brassica oleracea (wild cabbage)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T14532  
R;Kusaba, M.; Nishio, T.; Satta, Y.; Hinata, K.; Ockendon, D.  
Proc. Natl. Acad. Sci. U.S.A. 94, 7673-7678, 1997  
A;Title: Striking sequence similarity in inter- and intra-specific comparisons o f class  
echanism.  
A;Reference number: Z18078; MUID:97352858; PMID:9207151  
A;Accession: T14532  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-426 <KUS>  
A;Cross-references: UNIPROT:O23842; EMBL:D85208; NID:g2351147; PIDN:BAA21942.1; PID:g235  
C;Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology  
C;Keywords: glycoprotein

Query Match 39.0%; Score 41; DB 2; Length 426;  
Best Local Similarity 46.7%; Pred. No. 81;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 5 KFYEAANGTRDHHKG 19  
: | | | | : | | | |  
Db 202 EFYILVNGSRFHRSG 216

RESULT 23  
T14535  
S-locus-specific glycoprotein - wild cabbage (fragment)  
N;Alternate names: S glycoprotein  
C;Species: Brassica oleracea (wild cabbage)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T14535  
R;Kusaba, M.; Nishio, T.; Satta, Y.; Hinata, K.; Ockendon, D.  
Proc. Natl. Acad. Sci. U.S.A. 94, 7673-7678, 1997  
A;Title: Striking sequence similarity in inter- and intra-specific comparisons o f class  
echanism.  
A;Reference number: Z18078; MUID:97352858; PMID:9207151  
A;Accession: T14535  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-429 <KUS>  
A;Cross-references: UNIPROT:O23845; EMBL:D85211; NID:g2351153; PIDN:BAA21945.1; PID:g235  
C;Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology  
C;Keywords: glycoprotein

Query Match 39.0%; Score 41; DB 2; Length 429;  
Best Local Similarity 46.7%; Pred. No. 82;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 5 KFYEAANGTRDHHKG 19  
: | | | | : | | | |  
Db 202 EFYILVNGSRFHRSG 216

RESULT 24  
F82345  
conserved hypothetical protein VC0264 [imported] - Vibrio cholerae (strain N16961 serogr  
C;Species: Vibrio cholerae



C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C/Accession: F82345  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.K.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406: 477-483, 2000  
A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: F82345  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <HEI>  
A/Cross-references: UNIPROT:Q9KV92; GB:A5004114; GB:A5003852; NID:g9654662; PIDN:AAF93434  
A/Experimental source: serogroup O1, strain NI6961; biotype El Tor  
C/Genetics:  
A:Gene: VC0264  
A:Map position: 1

Query Match 39.0%; Score 41; DB 2; Length 455;  
Best Local Similarity 41.2%; Pred. No. 87;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKFVEAANGTRDHK 17  
|||::| : |||  
Db 202 RADYLFIETTYGDKQHE 218

RESULT 25  
A81732  
Glycogen synthase TC0181 [imported] - Chlamydia muridarum (strain Nigg)  
C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C/Accession: A81732  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A/Reference number: A81500; MUID:20150255; PMID:10684935  
A/Accession: A81732  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <TGT>  
A/Cross-references: UNIPROT:Q9PLC3; GB:A5002285; GB:A5002160; NID:g7190212; PIDN:AAF39059  
A/Experimental source: strain Nigg (MoPn)  
C/Genetics:  
A:Gene: TC0181  
C/Superfamily: starch synthase

Query Match 39.0%; Score 41; DB 2; Length 474;  
Best Local Similarity 38.9%; Pred. No. 90;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFVEAANGTRDHKG 19  
|||::| : |||  
Db 223 SDYEIHDAITARQHHLKG 240

RESULT 26  
E90086  
t-complex protein 1, delta SU [imported] - Guillardia theta nucleomorph  
C/Species: nucleomorph Guillardia theta  
A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C/Accession: E90086  
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reishus  
Nature 410, 1091-1096, 2001  
A>Title: The highly reduced  
A/Reference number: A99082; MUID:11323671; PMID:11323671  
A/Accession: E90086  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-519 <DOU>  
A/Cross-references: UNIPROT:Q9BRX6; GB:AFL65818; NID:gl3794449; PIDN:AAK39824.1; GSPDB:B



A;Reference number: JCS877; MUID:98014979; PMID:9353562  
A;Accession: JC5877  
A;Molecule type: mRNA  
A;Residues: 1-710 <YOS>  
A;Cross-references: UNIPROT.Q9XTA2; GB:AB028866; NID:g5103284; PIDN:BAA78907.1; PID:g5103284  
A;Experimental source: brain  
A;Note: in the authors' translation residues 124-129 and 130-172 are interchanged  
C;Comment: This proline specific endopeptidase cleaves the peptide bond at the carboxyl  
C;Superfamily: prolly oligopeptidase  
C;Keywords: hydrolase; serine proteinase  
F;554,642,680/Active site: Ser, Asp, His #status predicted

Query Match 38.1%; Score 40.5; DB 2; Length 710;  
Best Local Similarity 40.4%; Pred. No. 1.6e+02;  
Matches 11; Conservative 4; Mismatches 3; Indels 9; Gaps 2;

Qy 2 SDYK----FYEAANGTR-----DHKKG 19  
|||:::||:  
Db 433 SDYQTQVIFPSKDGTKIPMFIVHKKG 459  
|||:::||:

RESULT 34  
G86760  
diacylglycerol kinase (EC 2.7.1.107) [imported] - Lactococcus lactis subsp. lactis (strain ATCC 30564)  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: G86760  
R;Bolotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich, S.  
Genome Res. 11, 731-753, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. cremoris  
A;Reference number: AB6625; MUID:21235186; PMID:11337471  
A;Accession: G86760  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <STO>  
A;Cross-references: UNIPROT.Q9CGK9; GB:AE005176; PID:g12724043; PIDN:AAK05185.1; GSFPDB:G86760  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: dgka  
C;Superfamily: Bacillus subtilis diacylglycerol kinase dgka  
C;Keywords: phosphotransferase

Query Match 38.1%; Score 40; DB 2; Length 151;  
Best Local Similarity 44.4%; Pred. No. 42;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKFYEANGTRDHKKG 19  
|||:::||:  
Db 108 SDYQFYMRKRKRDMAAG 125  
|||:::||:

RESULT 35  
T16301  
hypothetical protein F38B6.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T16301  
R;Wilcox, L.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid F38B6.  
A;Reference number: Z18491  
A;Accession: T16301  
A;Status: preliminary; translated from GB/EMBL/DDBY  
A;Molecule type: DNA  
A;Residues: 1-218 <WIL>  
A;Cross-references: UNIPROT.Q20147; EMBL:U40060; NID:g1055174; PID:g1055180; PIDN:AAA811  
C;Genetics:  
A;Gene: CSP:F38B6.2  
A;Introns: 29/1; 54/1; 106/1; 135/3

Query Match 38.1%; Score 40; DB 2; Length 218;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Search completed: November 24, 2004, 09:30:12  
Job time : 17.6957 sec

```

RESULT 40
T43964
hypothetical protein U5 [imported] - human herpesvirus 6 (strain HST)
C:Species: human herpesvirus 6
A:Variety: strain HST
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43964
R:Inegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; Kawabuchi, T.;
J. Virol. 73, 8033-8063, 1999
A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and
A:Reference number: Z22732; MUID:59412319; PMID:10482554
A:Accession: T43964
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-443 <ISS>
A:Cross-reference: UNIPROT:O9WTS59; EMBL:AB021506; NID:G44995977; PTDN:BAA78225.1; PTD:G44995977

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 24, 2004, 09:29:11 ; Search time 71.663 Seconds  
(without alignments)  
94.040 Million cell updates/sec

Title: US-09-719-379A-1  
Perfect score: 105  
Sequence: 1 RSDYKFEANGTRDHKKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description       |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1          | 99    | 94.3        | 19     | US-10-203-942-1      | Sequence 1, Appli |
| 2          | 99    | 94.3        | 28     | US-10-203-942-5      | Sequence 5, Appli |
| 3          | 99    | 94.3        | 40     | US-10-223-711-10     | Sequence 10, Appl |
| 4          | 83    | 79.0        | 18     | US-10-223-711-1      | Sequence 1, Appli |
| 5          | 48    | 45.7        | 378    | US-10-739-930-5878   | Sequence 5878, Ap |
| 6          | 47    | 44.8        | 1226   | US-10-437-963-198352 | Sequence 198352,  |
| 7          | 46.5  | 44.3        | 114    | US-10-437-963-130405 | Sequence 130405,  |
| 8          | 46.5  | 44.3        | 345    | US-10-174-487-4      | Sequence 4, Appli |
| 9          | 46    | 43.8        | 478    | US-10-282-122A-69866 | Sequence 69866, A |
| 10         | 45    | 42.9        | 168    | US-10-739-930-10635  | Sequence 10635, A |
| 11         | 45    | 42.9        | 202    | US-10-359-012-19     | Sequence 19, Appl |
| 12         | 45    | 42.9        | 220    | US-10-425-115-255044 | Sequence 255044,  |
| 13         | 45    | 42.9        | 274    | US-10-437-963-142443 | Sequence 142443,  |

|    |      |      |      |    |                      |                   |
|----|------|------|------|----|----------------------|-------------------|
| 14 | 45   | 42.9 | 320  | 17 | US-10-739-930-10636  | Sequence 10636, A |
| 15 | 45   | 42.9 | 1245 | 16 | US-10-437-963-128610 | Sequence 128610,  |
| 16 | 45   | 42.9 | 2734 | 14 | US-10-359-012-12     | Sequence 12, Appl |
| 17 | 45   | 42.9 | 2932 | 16 | US-10-437-963-120244 | Sequence 120244,  |
| 18 | 45   | 42.9 | 3002 | 14 | US-10-359-012-10     | Sequence 10, Appl |
| 19 | 45   | 42.9 | 3002 | 14 | US-10-359-012-20     | Sequence 20, Appl |
| 20 | 45   | 42.9 | 4723 | 14 | US-10-359-013-8      | Sequence 8, Appli |
| 21 | 44.5 | 42.4 | 293  | 17 | US-10-425-115-205237 | Sequence 205237,  |
| 22 | 44   | 41.9 | 109  | 17 | US-10-425-115-204064 | Sequence 204064,  |
| 23 | 44   | 41.9 | 292  | 10 | US-09-844-948-4      | Sequence 4, Appli |
| 24 | 44   | 41.9 | 292  | 14 | US-10-184-648-22     | Sequence 22, Appl |
| 25 | 44   | 41.9 | 337  | 16 | US-10-767-701-39724  | Sequence 39724, A |
| 26 | 43.5 | 41.4 | 429  | 15 | US-10-389-566-1320   | Sequence 1320, Ap |
| 27 | 43.5 | 41.4 | 487  | 15 | US-10-425-114-64039  | Sequence 64039, A |
| 28 | 43.5 | 41.4 | 511  | 17 | US-10-425-115-225538 | Sequence 225538,  |
| 29 | 43   | 41.0 | 51   | 17 | US-10-425-115-185590 | Sequence 185590,  |
| 30 | 43   | 41.0 | 68   | 17 | US-10-425-115-225711 | Sequence 225711,  |
| 31 | 43   | 41.0 | 154  | 16 | US-10-767-701-50418  | Sequence 50418, A |
| 32 | 43   | 41.0 | 274  | 15 | US-10-424-599-160723 | Sequence 160723,  |
| 33 | 43   | 41.0 | 322  | 14 | US-10-177-980-13     | Sequence 13, Appl |
| 34 | 43   | 41.0 | 442  | 15 | US-10-424-599-159392 | Sequence 159392,  |
| 35 | 43   | 41.0 | 1299 | 16 | US-10-437-963-128293 | Sequence 128293,  |
| 36 | 42.5 | 40.5 | 24   | 10 | US-09-962-756-1781   | Sequence 1781, Ap |
| 37 | 42.5 | 40.5 | 24   | 14 | US-10-253-471-1781   | Sequence 1781, Ap |
| 38 | 42.5 | 40.5 | 24   | 15 | US-10-253-493-1781   | Sequence 1781, Ap |
| 39 | 42.5 | 40.5 | 75   | 15 | US-10-424-599-152876 | Sequence 152876,  |
| 40 | 42   | 40.0 | 13   | 14 | US-10-203-942-3      | Sequence 3, Appli |
| 41 | 42   | 40.0 | 22   | 14 | US-10-203-942-7      | Sequence 7, Appli |
| 42 | 42   | 40.0 | 66   | 15 | US-10-424-599-162635 | Sequence 162635,  |
| 43 | 42   | 40.0 | 467  | 15 | US-10-424-599-255389 | Sequence 255389,  |
| 44 | 42   | 40.0 | 1764 | 16 | US-10-437-963-132021 | Sequence 132021,  |
| 45 | 41.5 | 39.5 | 26   | 10 | US-09-962-756-1172   | Sequence 1172, Ap |

## ALIGNMENTS

RESULT 1  
US-10-203-942-1  
; Sequence 1, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOEL, PHILIPPE  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN AND USE THEREOF IN VACCINATION  
; FILE REFERENCE: B45210  
; CURRENT APPLICATION NUMBER: US/10/203,942  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-1

Query Match 94.3%; Score 99; DB 14; Length 19;  
Best Local Similarity 94.7%; Pred. No. 5.1e-09;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFEANGTRDHKKG 19  
|||||  
Db 1 RSDYKFEANGTRDHKKG 19

RESULT 2



```

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 198352
; LENGTH: 1226
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAV
US-10-437-963-198352

```

Query Match 44.8%; Score 47; DB 16; Length 1226;  
Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels

Qy 1 RSDYKFYEAAANGTRDHK 17  
: | : | | : | |  
Db 1047 ODDWSFEEAYNLSRDHK 1063

RESULT 7  
US-10-437-963-130405  
; Sequence 130405. Application US/10437963

|                          |       |               |           |             |
|--------------------------|-------|---------------|-----------|-------------|
| Query Match              | 44.3% | Score 46.5;   | DB 16;    | Length 114; |
| Best Local Similarity    | 46.4% | Pred. NO. 12; |           |             |
| Matches 13; Conservative |       | 1; Mismatches | 5; Indels | 9           |

**Qy** 1 RSDYKFEAA---NG-----TRDHKKG 19  
| | : | | | |  
**Db** 62 REGYRFSTAATVYNNGAQIGPTNDHKKG 89

RESULT 8  
US-10-174-487-4  
; Sequence 4, Application US/10174487  
; Publication No. US20040106173A1  
; GENERAL INFORMATION:  
; APPLICANT: Verde, Fulvia  
; APPLICANT: Catanuto, Paola  
; APPLICANT: Wiley, David J.  
; APPLICANT: You, Min

```

; PRIOR FILING DATE: 2001-06-19
;
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1.1
; SEQ ID NO 4
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-174-487-4

```

Query Match 44.3%; Score 46.5; DB 16; Length 345;  
Best Local Similarity 40.7%; Pred. No. 38;  
Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Q7 1 RSDYKF-----YEAANGTRDHKK 18  
||| ||| | : ||| ||  
D6 304 RSOYKFETNAKVGKVGRYRGSGNRDNKK 330

RESULT 9  
US-10-282-122A-69866  
; Sequence 69866, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

Query Match 43.8%; Score 46; DB 15; Length 478;  
Best Local Similarity 60.0%; Pred. No. 64;  
Matches 9: Conservative 1; Mismatches 5 Indels

Qv 2 SDYKFYEANGTRDH 16

Db 63 SSWKPYEAKNGFAVH 77  
| : ||||| || |

## RESULT 10

US-10-739-930-10635  
; Sequence 10635, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 10635  
; LENGTH: 168  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(168)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C5473\_1.p  
US-10-739-930-10635

Query Match 42.9%; Score 45; DB 17; Length 168;  
Best Local Similarity 42.1%; Pred. No. 31;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKPYEAAANGTRDHKKG 19  
: : ||||| || |  
Db 12 QGDSRFYDAARARRGHHG 30

## RESULT 11

US-10-359-012-19  
; Sequence 19, Application US/10359012  
; Publication No. US2003023418A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Drosophila truncated mutant  
US-10-359-012-19

Query Match 42.9%; Score 45; DB 14; Length 202;  
Best Local Similarity 61.1%; Pred. No. 38;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKPYEAAANGTRDHKK 18  
||| : | | | | |  
Db 5 RSDDESYP--ANETREHKK 20

## RESULT 12

US-10-425-115-255044  
; Sequence 255044, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 255044  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_164177C.1.p  
US-10-425-115-255044

Query Match 42.9%; Score 45; DB 17; Length 220;  
Best Local Similarity 42.1%; Pred. No. 41;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKPYEAAANGTRDHKKG 19  
: : ||||| || |  
Db 12 QGDSRFYDAARARRGHHG 30

## RESULT 13

US-10-437-963-142443  
; Sequence 142443, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 142443  
; LENGTH: 274  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_43449C.1.p  
US-10-437-963-142443

Query Match 42.9%; Score 45; DB 16; Length 274;  
Best Local Similarity 61.5%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKPYEAAANGTR 14  
| | | | |  
Db 90 STYSFHDAAANGNR 102



## RESULT 14

US-10-739-930-10636  
; Sequence 10636, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: KOVALIC, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 10636  
; LENGTH: 320  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: TRIAB-23APR03-C5473\_3.p  
US-10-739-930-10636

Query Match 42.9%; Score 45; DB 17; Length 320;  
Best Local Similarity 42.1%; Pred. No. 61;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKPYEAANGTRDHKKG 19  
Db 12 QGDSRFYDAAPARRGHHG 30

## RESULT 15

US-10-437-963-128610  
; Sequence 128610, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: KOVALIC, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 128610  
; LENGTH: 1245  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(1245)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_30949C.1.p  
US-10-437-963-128610

Query Match 42.9%; Score 45; DB 16; Length 1245;  
Best Local Similarity 61.5%; Pred. No. 2.5e+02;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 YEAAANGTRDHKKG 19  
Db 617 YEKQGGGRDHNGK 629

## RESULT 16

US-10-359-012-12  
; Sequence 12, Application US/10359012  
; Publication No. US2003023419A1

## ; GENERAL INFORMATION:

; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; TITLE OF INVENTION: AND METHODS OF USING THE SAME  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 2734  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-10-359-012-12

Query Match 42.9%; Score 45; DB 14; Length 2734;  
Best Local Similarity 61.1%; Pred. No. 5.8e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKPYEAANGTRDHKK 18  
Db 2537 RSDDESY--ANETREHKK 2552

## RESULT 17

US-10-437-963-120244  
; Sequence 120244, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: KOVALIC, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 120244  
; LENGTH: 2932  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_23383C.1.p  
US-10-437-963-120244

Query Match 42.9%; Score 45; DB 16; Length 2932;  
Best Local Similarity 61.5%; Pred. No. 6.3e+02;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 YEAAANGTRDHKKG 19  
Db 416 YERQGGGRDHNGK 428

## RESULT 18

US-10-359-012-10

; Sequence 10, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; TITLE OF INVENTION: AND METHODS OF USING THE SAME  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; PRIOR FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 3002  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-10-359-012-10

Query Match 42.9%; Score 45; DB 14; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 6.4e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEANGTRDHKK 18  
||| : | ||| : |||  
Db 2805 RSDDESY--ANETREHKK 2820

RESULT 19  
US-10-359-012-20  
; Sequence 20, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; TITLE OF INVENTION: AND METHODS OF USING THE SAME  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; PRIOR FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3002  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Drosophila G-W mutant. G residues 134, 136, 139 of Drosophila M  
US-10-359-012-20

Query Match 42.9%; Score 45; DB 14; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 6.4e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEANGTRDHKK 18  
||| : | ||| : |||  
Db 2805 RSDDESY--ANETREHKK 2820

RESULT 20  
US-10-359-012-8  
; Sequence 8, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; TITLE OF INVENTION: AND METHODS OF USING THE SAME  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; PRIOR FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 4723  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-10-359-012-8

Query Match 42.9%; Score 45; DB 14; Length 4723;  
Best Local Similarity 61.1%; Pred. No. 1e+03;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEANGTRDHKK 18  
||| : | ||| : |||  
Db 4526 RSDDESY--ANETREHKK 4541

RESULT 21  
US-10-425-115-205237  
; Sequence 205237, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 205237  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_118765C.1.pep  
US-10-425-115-205237

Query Match 42.4%; Score 44.5; DB 17; Length 293;  
Best Local Similarity 76.9%; Pred. No. 67;  
Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 2 SDY-KFYEANGT 13

```
Db      126 SDYDFEYDANGT 138
||||:|||||
RESULT 22
US-10-425-115-204064
; Sequence 204064, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 204064
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_117695C.1.pep
US-10-425-115-204064
Query Match      41.9%; Score 44; DB 17; Length 109;
Best Local Similarity 47.4%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Qy      1 RSDYKFEYEAANGTRDHHKG 19
|:|||||:|||||
Db      21 RALYKVERGAAGQEHKKG 39
|||||:|||||
RESULT 23
US-09-844-948-4
; Sequence 4, Application US/09844948
; Publication No. US20030119161A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: 32132, A NOVEL FUCOSYLTRANSFERASE FAMILY
; FILE REFERENCE: 10448-048001
; CURRENT APPLICATION NUMBER: US/09/844,948
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: US 60/200,604
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-844-948-4
Query Match      41.9%; Score 44; DB 10; Length 292;
Best Local Similarity 53.3%; Pred. No. 81;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy      2 SDYKFEYEAANGTRDH 16
|||||:|||||
Db      206 SKYKFLAFENSKDH 220
|||||:|||||
RESULT 24
US-10-184-648-22
; Sequence 22, Application US/10184648
; Publication No. US20030224376A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Williamson, Mark
; APPLICANT: Tsai, Fong-Ying
; APPLICANT: Hunter, John J.
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Olandt, Peter J.
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-192001
; CURRENT APPLICATION NUMBER: US/10/184,648
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: US 09/815,028
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: PCT/US01/09358
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,964
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/801,220
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/US01/07269
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/187,456
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/816,714
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: PCT/US01/09468
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,865
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/844,948
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/13805
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/200,604
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 09/861,164
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16292
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,408
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/883,060
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19138
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,079
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/962,678
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: PCT/US01/29963
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/235,044
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 09/973,457
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/238,849
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 10/072,285
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: PCT/US02/03736
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/267,494
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: US 09/817,910
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: PCT/US01/09633
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,092
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/842,528
```

; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: PCT/US01/40607  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: US 60/199,500  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: US 09/882,836  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: PCT/US01/19543  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/211,730  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 09/882,872  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: PCT/US01/19153  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/212,077  
; PRIOR FILING DATE: 2000-06-15  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: consensus sequence  
US-10-184-648-22

Query Match 41.9%; Score 44; DB 14; Length 292;  
Best Local Similarity 53.3%; Pred. No. 81;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKFEAANGTRDH 16  
||||| :||  
Db 206 SKYFYLAFENSKDH 220

## RESULT 25

US-10-767-701-39724  
; Sequence 39724, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53515)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 39724  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C28861\_1.pep  
US-10-767-701-39724

Query Match 41.9%; Score 44; DB 16; Length 337;  
Best Local Similarity 53.3%; Pred. No. 94;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKFEAANGTRD 15  
||||| :||  
Db 274 RSDQALYDGGNGTTD 288

## RESULT 26

US-10-389-566-1320  
; Sequence 1320, Application US/10389566  
; Publication No. US20040025202A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Laurie, Cathy C  
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants  
; FILE REFERENCE: 38-77(52900)D  
; CURRENT APPLICATION NUMBER: US/10/389,566  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/365,301  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/391,786  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/392,018  
; PRIOR FILING DATE: 2002-06-26  
; NUMBER OF SEQ ID NOS: 2459  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1320  
; LENGTH: 429  
; TYPE: PRT  
; ORGANISM: Sulfolobus tokodaii  
US-10-389-566-1320

Query Match 41.4%; Score 43.5; DB 15; Length 429;  
Best Local Similarity 56.2%; Pred. No. 1.5e+02;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 4 YKFEAANG-TRDHKK 18  
||||| :||  
Db 348 YKFEIVSGSTRNHR 363

## RESULT 27

US-10-425-114-64039  
; Sequence 64039, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 64039  
; LENGTH: 487  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3245-500-B3\_FLI.pep  
US-10-425-114-64039

Query Match 41.4%; Score 43.5; DB 15; Length 487;  
Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
Matches 10; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 7 YEAANG-TRDHKKG 19  
||||| :||  
Db 269 YEAPHGLTRDHKG 282

## RESULT 28

US-10-425-115-225538  
; Sequence 225538, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants

FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 225538  
LENGTH: 511  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_137279C.1.pep  
US-10-425-115-225538

Query Match 41.4%; Score 43.5; DB 17; Length 511;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 10; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 7 YEAAAG-TRDHKKG 19  
Db 293 YEAPHGLTRDHKG 306

## RESULT 29

US-10-425-115-185590  
Sequence 185590, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 185590  
LENGTH: 51  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_100847C.1.pep  
US-10-425-115-185590

Query Match 41.0%; Score 43; DB 17; Length 51;  
Best Local Similarity 38.9%; Pred. No. 19;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFEAANGTRDHKK 18  
Db 26 RSSRYTGAASSVGVHER 43

## RESULT 30

US-10-425-115-225711  
Sequence 225711, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 225711  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:

Query Match 41.0%; Score 43; DB 15; Length 274;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OTHER INFORMATION: Clone ID: MRT4577\_137436C.1.pep  
US-10-425-115-225711

Query Match 41.0%; Score 43; DB 17; Length 68;  
Best Local Similarity 41.2%; Pred. No. 25;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKFEAANGTRDHKK 18  
Db 52 SDTSYIRDSRDRHQ 68

## RESULT 31

US-10-767-701-50418  
Sequence 50418, Application US/10767701  
Publication No. US20040172684A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53535)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 50418  
LENGTH: 154  
TYPE: PRT  
ORGANISM: Sorghum bicolor  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1).(154)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3480-006-PI-K1-C5.pep  
US-10-767-701-50418

Query Match 41.0%; Score 43; DB 16; Length 154;  
Best Local Similarity 52.6%; Pred. No. 60;  
Matches 10; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

Qy 3 DYKFEAANGTRDHKKG 19  
Db 103 DYDFYXEEKANARRAEKRG 121

## RESULT 32

US-10-424-599-160723  
Sequence 160723, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 160723  
LENGTH: 274  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_116150C.1.pep  
US-10-424-599-160723

Query Match 41.0%; Score 43; DB 15; Length 274;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 YKFEAANGTRDHHKG 19  
:| : :|||  
Db 172 HKSCQKEDGTDHKG 187

## RESULT 33

US-10-177-980-13  
; Sequence 13, Application US/10177980  
; Publication No. US20030166232A1  
; GENERAL INFORMATION:  
; APPLICANT: Saraz, Jan  
; APPLICANT: Franz, Petra  
; APPLICANT: Aspenstrm, Pontus  
; APPLICANT: Hellman, Ulf  
; APPLICANT: Gonez, Leonel Jorge  
; APPLICANT: Hedin, Carl-Henrik  
; TITLE OF INVENTION: PARF, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1  
; FILE REFERENCE: L0461/7030  
; CURRENT APPLICATION NUMBER: US/10/177,980  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US/09/080,855  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 08/805,583  
; PRIOR FILING DATE: 1997-02-25  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-980-13

Query Match 41.0%; Score 43; DB 14; Length 322;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 FYEAANGTRDHHKG 19  
:| : :|||  
Db 250 FHLSNATRDYQPG 263

## RESULT 34

US-10-424-599-159392  
; Sequence 159392, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 159392  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(442)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_114951C.1.pep  
US-10-424-599-159392

Query Match 41.0%; Score 43; DB 15; Length 442;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 YKFEAANGTRDHHKG 19  
:| : :|||  
Db 175 HKSCQKEDGTDHKG 190

## RESULT 35

US-10-437-963-128293  
; Sequence 128293, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 128293  
; LENGTH: 1299  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_30660C.1.pep  
US-10-437-963-128293

Query Match 41.0%; Score 43; DB 16; Length 1299;  
Best Local Similarity 50.0%; Pred. No. 5.6e+02;  
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 SDYKFEAANGTRDHHKG 19  
:| : :|||  
Db 123 NDKKYEKQGERVHNGK 140

## RESULT 36

US-09-962-756-1781  
; Sequence 1781, Application US/09962756  
; Publication No. US20030195147A1  
; GENERAL INFORMATION:  
; APPLICANT: PILLUTLA, RENUKA  
; APPLICANT: BRISSETTE, RENEE  
; APPLICANT: BLUME, ARTHUR J.  
; APPLICANT: SCHAFER, LAUGE  
; APPLICANT: BRANDT, JAKOB  
; APPLICANT: GOLDSTEIN, NEIL I.  
; APPLICANT: SPETZLER, JANE  
; APPLICANT: OSTERGAARD, SOREN  
; APPLICANT: HANSEN, PER HERTZ  
; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
; FILE REFERENCE: 1878-4051US1  
; CURRENT APPLICATION NUMBER: US/09/962,756  
; CURRENT FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 09/538,038  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 09/146,127  
; PRIOR FILING DATE: 1998-09-02  
; NUMBER OF SEQ ID NOS: 2227  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1781  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-962-756-1781

```
Query Match      40.5%; Score 42.5; DB 10; Length 24;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 12; Conservative 1; Mismatches 2; Indels 9; Gaps 2;

Qy 3 DYK-FYEA-----ANGTRDHK 17
    ||| |||:| |||||
Db 1 DYKDFYDAIQIVRSARAGGTRDKK 24

RESULT 37
US-10-253-471-1781
; Sequence 1781, Application US/10253471
; Publication No. US20030236190A1
; GENERAL INFORMATION:
; APPLICANT: PILUTULA, RENUKA et al.
; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS
; FILE REFERENCE: 1878-4057
; CURRENT APPLICATION NUMBER: US/10/253,471
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: 09/962,756
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/538,038
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 09/146,127
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-253-471-1781

Query Match      40.5%; Score 42.5; DB 14; Length 24;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 12; Conservative 1; Mismatches 2; Indels 9; Gaps 2;

Qy 3 DYK-FYEA-----ANGTRDHK 17
    ||| |||:| |||||
Db 1 DYKDFYDAIQIVRSARAGGTRDKK 24

RESULT 38
US-10-253-493-1781
; Sequence 1781, Application US/10253493
; Publication No. US20040023887A1
; GENERAL INFORMATION:
; APPLICANT: PILUTULA, RENUKA et al.
; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS
; FILE REFERENCE: 1878-4056
; CURRENT APPLICATION NUMBER: US/10/253,493
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: 09/962,756
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/538,038
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 09/146,127
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-253-493-1781

Query Match      40.5%; Score 42.5; DB 15; Length 24;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 12; Conservative 1; Mismatches 2; Indels 9; Gaps 2;

Qy 3 DYK-FYEA-----ANGTRDHK 17
    ||| |||:| |||||
Db 1 DYKDFYDAIQIVRSARAGGTRDKK 24

RESULT 39
US-10-424-599-152876
; Sequence 152876, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 152876
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109070C.1.pap
US-10-424-599-152876

Query Match      40.5%; Score 42.5; DB 15; Length 75;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 2;

Qy 1 RSDYKFP-YEANGTRDHHKG 19
    : |||:| || |||||
Db 8 KGDYKGYSEKNG--DHKHG 25

RESULT 40
US-10-203-942-3
; Sequence 3, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows, Version 4.0
; SEQ ID NO 3
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-3

Query Match      40.0%; Score 42; DB 14; Length 13;
Best Local Similarity 87.5%; Pred. No. 6.5;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYE 8
    |||||
```

Db 1 RSDYKFYD 8

Search completed: November 24, 2004, 10:00:11  
Job time : 72.663 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 24, 2004, 09:29:11 ; Search time 71.663 Seconds  
(without alignments)  
94.040 Million cell updates/sec

Title: US-09-719-379A-5  
Perfect score: 107  
Sequence: 1 RSDYKFDANGTRDHKKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PTCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PTCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description       |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1          | 107   | 100.0       | 19     | US-10-203-942-1      | Sequence 1, Appli |
| 2          | 107   | 100.0       | 28     | US-10-203-942-5      | Sequence 5, Appli |
| 3          | 107   | 100.0       | 40     | US-10-223-711-10     | Sequence 10, Appl |
| 4          | 91    | 85.0        | 18     | US-10-223-711-1      | Sequence 1, Appli |
| 5          | 52.5  | 49.1        | 293    | US-10-425-115-205237 | Sequence 205237,  |
| 6          | 49.5  | 46.3        | 326    | US-10-767-701-44199  | Sequence 44199, A |
| 7          | 49    | 45.8        | 99     | US-10-767-701-51811  | Sequence 51811, A |
| 8          | 49    | 45.8        | 401    | US-10-146-473-76     | Sequence 76, Appl |
| 9          | 47    | 43.9        | 378    | US-10-739-930-5878   | Sequence 5878, Ap |
| 10         | 47    | 43.9        | 1561   | US-10-282-122A-72279 | Sequence 72279, A |
| 11         | 45.5  | 42.5        | 75     | US-10-424-599-152876 | Sequence 152876,  |
| 12         | 45.5  | 42.5        | 345    | US-10-174-487-4      | Sequence 4, Appli |
| 13         | 45    | 42.1        | 202    | US-10-359-012-19     | Sequence 19, Appl |

|    |      |      |      |    |                      |                    |
|----|------|------|------|----|----------------------|--------------------|
| 14 | 45   | 42.1 | 824  | 15 | US-10-425-114-68661  | Sequence 68661, A  |
| 15 | 45   | 42.1 | 867  | 17 | US-10-425-115-287968 | Sequence 287968, A |
| 16 | 45   | 42.1 | 972  | 15 | US-10-425-114-65128  | Sequence 65128, A  |
| 17 | 45   | 42.1 | 1118 | 17 | US-10-425-115-287946 | Sequence 287946,   |
| 18 | 45   | 42.1 | 1245 | 16 | US-10-437-963-128610 | Sequence 128610,   |
| 19 | 45   | 42.1 | 2734 | 14 | US-10-359-012-12     | Sequence 12, Appl  |
| 20 | 45   | 42.1 | 3002 | 14 | US-10-359-012-20     | Sequence 10, Appl  |
| 21 | 45   | 42.1 | 3002 | 14 | US-10-359-012-20     | Sequence 20, Appl  |
| 22 | 45   | 42.1 | 4723 | 14 | US-10-359-012-8      | Sequence 8, Appli  |
| 23 | 44   | 41.1 | 399  | 15 | US-10-424-599-170099 | Sequence 170099,   |
| 24 | 44   | 41.1 | 643  | 16 | US-10-437-963-113950 | Sequence 113950,   |
| 25 | 44   | 41.1 | 2932 | 16 | US-10-437-963-120244 | Sequence 120244,   |
| 26 | 43.5 | 40.7 | 613  | 17 | US-10-425-115-283230 | Sequence 283230,   |
| 27 | 43.5 | 40.7 | 1001 | 16 | US-10-437-963-143107 | Sequence 143107,   |
| 28 | 43   | 40.2 | 13   | 14 | US-10-203-942-3      | Sequence 3, Appli  |
| 29 | 43   | 40.2 | 22   | 14 | US-10-203-942-7      | Sequence 7, Appli  |
| 30 | 43   | 40.2 | 109  | 17 | US-10-425-115-204064 | Sequence 204064,   |
| 31 | 43   | 40.2 | 154  | 16 | US-10-767-701-50418  | Sequence 50418, A  |
| 32 | 43   | 40.2 | 274  | 15 | US-10-424-599-160723 | Sequence 160723,   |
| 33 | 43   | 40.2 | 337  | 16 | US-10-767-701-39724  | Sequence 39724, A  |
| 34 | 43   | 40.2 | 1299 | 16 | US-10-437-963-128293 | Sequence 128293,   |
| 35 | 42.5 | 39.7 | 81   | 15 | US-10-424-599-162795 | Sequence 162795,   |
| 36 | 42.5 | 39.7 | 202  | 15 | US-10-424-599-285321 | Sequence 285321,   |
| 37 | 42.5 | 39.7 | 995  | 14 | US-10-369-493-6568   | Sequence 6568, Ap  |
| 38 | 42   | 39.3 | 91   | 15 | US-10-424-599-272502 | Sequence 272502,   |
| 39 | 42   | 39.3 | 189  | 15 | US-10-289-762-135    | Sequence 135, App  |
| 40 | 42   | 39.3 | 361  | 15 | US-10-282-122A-54860 | Sequence 54860, A  |
| 41 | 42   | 39.3 | 407  | 14 | US-10-369-493-19648  | Sequence 19648, A  |
| 42 | 42   | 39.3 | 442  | 15 | US-10-424-599-159392 | Sequence 159392,   |
| 43 | 42   | 39.3 | 448  | 15 | US-10-108-260A-4755  | Sequence 4755, Ap  |
| 44 | 42   | 39.3 | 471  | 15 | US-10-108-260A-2588  | Sequence 2588, Ap  |
| 45 | 42   | 39.3 | 627  | 14 | US-10-153-668-136    | Sequence 136, App  |

## ALIGNMENTS

RESULT 1  
US-10-203-942-1  
; Sequence 1, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOEL, PHILIPPE  
; APPLICANT: POOLMAN, JAN  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; FILE REFERENCE: B45210  
; CURRENT APPLICATION NUMBER: US/10/203,942  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-1

Query Match 100.0%; Score 107; DB 14; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDANGTRDHKKG 19  
|  
Db 1 RSDYKFDANGTRDHKKG 19

RESULT 2

US-10-203-942-5  
; Sequence 5, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOEL, PHILIPPE  
; APPLICANT: POULMAN, JAN  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN AND USE THEREOF IN VACCINATION  
; FILE REFERENCE: B45210  
; CURRENT APPLICATION NUMBER: US/10/203,942  
; PRIOR FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-5

Query Match 100.0%; Score 107; DB 14; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19  
Db 3 RSDYKFYEDANGTRDHKKG 21

RESULT 3  
US-10-223-711-10  
; Sequence 10, Application US/10223711  
; Publication No. US20030113344A1  
; GENERAL INFORMATION:  
; APPLICANT: Bakaletz, Lauren O.  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; FILE REFERENCE: 18525/04058  
; CURRENT APPLICATION NUMBER: US/10/223,711  
; CURRENT FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: 09/148,711  
; PRIOR FILING DATE: 1998-09-04  
; PRIOR APPLICATION NUMBER: 08/460,502  
; PRIOR FILING DATE: 1995-06-02 ✓  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: synthetic construct  
US-10-223-711-10

Query Match 100.0%; Score 107; DB 14; Length 40;  
Best Local Similarity 100.0%; Pred. No. 4e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19  
Db 1 RSDYKFYEDANGTRDHKKG 19

RESULT 4  
US-10-223-711-1  
; Sequence 1, Application US/10223711  
; Publication No. US20030113344A1  
; GENERAL INFORMATION:

; APPLICANT: Bakaletz, Lauren O.  
; APPLICANT: Kaumaya, Pravin T.P.  
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides  
; FILE REFERENCE: 18525/04058  
; CURRENT APPLICATION NUMBER: US/10/223,711  
; CURRENT FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: 09/148,711  
; PRIOR FILING DATE: 1998-09-04  
; PRIOR APPLICATION NUMBER: 08/460,502  
; PRIOR FILING DATE: 1995-06-02  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-223-711-1

Query Match 85.0%; Score 91; DB 14; Length 18;  
Best Local Similarity 88.9%; Pred. No. 5.2e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKK 18  
Db 1 RSDYKFYEDLNGTRNHKK 18

RESULT 5  
US-10-425-115-205237  
; Sequence 205237, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 205237  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_118765C.1.pep  
US-10-425-115-205237

Query Match 49.1%; Score 52.5; DB 17; Length 293;  
Best Local Similarity 84.6%; Pred. No. 7.4;  
Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 2 SDY-KFYEDANGT 13  
Db 126 SDYDEFYEDANGT 138

RESULT 6  
US-10-767-701-44199  
; Sequence 44199, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128

; SEQ ID NO 44199  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C43733\_1.pap  
US-10-767-701-44199

Query Match 46.3%; Score 49.5; DB 16; Length 326;  
Best Local Similarity 76.9%; Pred. No. 24;  
Matches 10; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 2 SDY-KFYEDANGT 13  
||| :|||  
Db 162 SDYFYEDANGT 174

## RESULT 7

US-10-767-701-51811  
; Sequence 51811, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 51811  
; LENGTH: 99  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB5048-010-R1-XPI-E2.pap  
US-10-767-701-51811

Query Match 45.8%; Score 49; DB 16; Length 99;  
Best Local Similarity 44.4%; Pred. No. 8.4;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSDYKFYEDANGTRDHKK 18  
|:|:|:|:|:|:|  
Db 75 RTFYEVAVDSGGTDRHK 92

## RESULT 8

US-10-146-473-76  
; Sequence 76, Application US/10146473  
; Publication No. US20030108889A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew  
; APPLICANT: Gout, Ivan  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Gure, Ali  
; APPLICANT: Chen, Yao-Tseng  
; APPLICANT: Old, Lloyd  
; TITLE OF INVENTION: Breast Cancer Antigens  
; FILE REFERENCE: L00461/70130(JRV)  
; CURRENT APPLICATION NUMBER: US/10/146,473  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 76  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE

; LOCATION: (292)..(292)  
; OTHER INFORMATION: X = any amino acid  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (355)..(355)  
; OTHER INFORMATION: X = any amino acid  
US-10-146-473-76

Query Match 45.8%; Score 49; DB 14; Length 401;  
Best Local Similarity 60.0%; Pred. No. 35;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 SDYKFYEDANGTRDH 16  
|||:|:|:|:|  
Db 207 SDYRFLEDVARTADH 221

## RESULT 9

US-10-739-930-5878  
; Sequence 5878, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 5878  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C17092\_1.p  
US-10-739-930-5878

Query Match 43.9%; Score 47; DB 17; Length 378;  
Best Local Similarity 50.0%; Pred. No. 67;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKFYEDANGTRDHKKG 19  
|||:|:|:|:|  
Db 186 SDEKLYKGINGYTDHKAG 203

## RESULT 10

US-10-282-122A-72279  
; Sequence 72279, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26

```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72279
; LENGTH: 1561
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (90)..(90)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-72279

Query Match          43.9%; Score 47; DB 15; Length 1561;
Best Local Similarity 47.1%; Pred. No. 2.8e+02;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy      3 DYKFVEDANGTRDHKKG 19
Db      1377 EYNFYDDYDQGDHYTG 1393

RESULT 11
US-10-424-599-152876
; Sequence 152876, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 152876
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109070C.1.pep
US-10-424-599-152876

Query Match          42.5%; Score 45.5; DB 15; Length 75;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 10; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Qy      1 RSDYKF-YEDANGTRDHKKG 19
Db      8 KGDYKYGSEKNG--DHKHG 25

RESULT 12
US-10-174-487-4
; Sequence 4, Application US/10174487
; Publication No. US20040106173A1
; GENERAL INFORMATION:
; APPLICANT: Verde, Fulvia
; APPLICANT: Catanuto, Paola
; APPLICANT: Wiley, David J.
; APPLICANT: You, Min
; TITLE OF INVENTION: BOT1: Target for Antifungal Agents
; FILE REFERENCE: 4052-24
; CURRENT APPLICATION NUMBER: US/10/174,487
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/298,901
; PRIOR FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-174-487-4

Query Match          42.5%; Score 45.5; DB 16; Length 345;
Best Local Similarity 40.7%; Pred. No. 1e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Qy      1 RSDYKF-----YEDANGTRDHKK 18
Db      304 RSQYKFTNAKVGKVGRYGSGNRDNKK 330

RESULT 13
US-10-359-012-19
; Sequence 19, Application US/10359012
; Publication No. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT
; TITLE OF INVENTION: AND METHODS OF USING THE SAME
; FILE REFERENCE: JHU1840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/388,325
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/384,302
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/354,178
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Drosophila truncated mutant
US-10-359-012-19

Query Match          42.1%; Score 45; DB 14; Length 202;
Best Local Similarity 61.1%; Pred. No. 72;
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy      1 RSDYKFYEDANGTRDHKK 18
Db      5 RSDDSY--ANETREHKK 20

RESULT 14
US-10-425-114-68661
; Sequence 68661, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68661
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73046D10_FLI.pep
US-10-425-114-68661

Query Match      42.1%; Score 45; DB 15; Length 824;
Best Local Similarity 47.1%; Pred. No. 3e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      3 DYKFYEDANGTRDHKKG 19
Db      70 DNKYHADKHGLQDAKKG 86

RESULT 15
US-10-425-115-287968
; Sequence 287968, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 287968
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(867)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_2571C.1.pep
US-10-425-115-287968

Query Match      42.1%; Score 45; DB 17; Length 867;
Best Local Similarity 47.1%; Pred. No. 3.2e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      3 DYKFYEDANGTRDHKKG 19
Db      113 DNKYHADKHGLQDAKKG 129

RESULT 16
US-10-425-114-65128
; Sequence 65128, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65128
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17067E07_FLI.pep
US-10-425-114-65128

Query Match      42.1%; Score 45; DB 15; Length 972;
Best Local Similarity 47.1%; Pred. No. 3.6e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      3 DYKFYEDANGTRDHKKG 19
Db      218 DNKYHADKHGLQDAKKG 234

RESULT 17
US-10-425-115-287946
; Sequence 287946, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 287946
; LENGTH: 1118
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_2569C.1.pep
US-10-425-115-287946

Query Match      42.1%; Score 45; DB 17; Length 1118;
Best Local Similarity 47.1%; Pred. No. 4.1e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      3 DYKFYEDANGTRDHKKG 19
Db      364 DNKYHADKHGLQDAKKG 380

RESULT 18
US-10-437-963-128610
; Sequence 128610, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 128610  
; LENGTH: 1245  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1245)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_30949C.1.pep  
US-10-437-963-128610

Query Match 42.1%; Score 45; DB 16; Length 1245;  
Best Local Similarity 61.5%; Pred. No. 4.6e+02;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 YEDANGTRDHKKG 19  
||| : ||| : ||| :  
Db 617 YEKQGERDHNGK 629

RESULT 19  
US-10-359-012-12  
; Sequence 12, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 2734  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-10-359-012-12

Query Match 42.1%; Score 45; DB 14; Length 2734;  
Best Local Similarity 61.1%; Pred. No. 1e+03;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEDANGTRDHKK 18  
||| : ||| : ||| :  
Db 2537 RSDDESJ--ANETREHKK 2552

RESULT 20  
US-10-359-012-10  
; Sequence 10, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi

; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 3002  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-10-359-012-10

Query Match 42.1%; Score 45; DB 14; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 1.1e+03;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEDANGTRDHKK 18  
||| : ||| : ||| :  
Db 2805 RSDDESJ--ANETREHKK 2820

RESULT 21  
US-10-359-012-20  
; Sequence 20, Application US/10359012  
; Publication No. US20030232419A1  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: KOLODKIN, Alex L.  
; APPLICANT: TERMAN, Jon R.  
; APPLICANT: MAO, Tianyi  
; APPLICANT: PASTERKAMP, Ronald J.  
; APPLICANT: YU, Hung-Hsiang  
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEPT  
; FILE REFERENCE: JHU1840-3  
; CURRENT APPLICATION NUMBER: US/10/359,012  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/388,325  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/384,302  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: US 60/354,178  
; PRIOR FILING DATE: 2002-02-04  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3002  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Drosophila G-W mutant. G residues 134, 136, 139 of Drosophila MI  
; OTHER INFORMATION: CAL changed to W residues  
US-10-359-012-20

Query Match 42.1%; Score 45; DB 14; Length 3002;  
Best Local Similarity 61.1%; Pred. No. 1.1e+03;  
Matches 11; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 1 RSDYKFYEDANGTRDHKK 18  
||| : ||| : ||| :  
Db 2805 RSDDESJ--ANETREHKK 2820

RESULT 22

```

RESULT 24
US-10-437-963-113950
; Sequence 113950, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113950
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_1768C.1.pap
US-10-437-963-113950

Query Match 41.1%; Score 44; DB 16; Length 643;
Best Local Similarity 47.1%; Pred. No. 3.3e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DYKPYEDANGTRDHKKG 19
|: |||: | |||
Db 587 DFNFYERKESARCAKKG 603

RESULT 25
US-10-437-963-120244
; Sequence 120244, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 120244
; LENGTH: 2932
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_23383C.1.pap
US-10-437-963-120244

Query Match 41.1%; Score 44; DB 16; Length 2932;
Best Local Similarity 61.5%; Pred. No. 1.6e+03;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 YEDANGTRDHKKG 19
| | | | | | |
Db 416 YEROGGERDHKKG 428

```

```

; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN AND USE THEREOF IN VACCINATION
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-3

Query Match          40.2%; Score 43; DB 14; Length 13
Best Local Similarity 77.8%; Pred. No. 8.8;
Matches 7; Conservative 2; Mismatches 0; Indels

Qy 1 RSDYKEYED 9
    |||||::
Db 1 RSDYKEYDN 9

RESULT 29
US-10-203-942-7
; Sequence 7, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN AND USE THEREOF IN VACCINATION
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-7

Query Match          40.2%; Score 43; DB 14; Length 22
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 0; Indels

Qy 1 RSDYKEYED 9
    |||||::
Db 3 RSDYKEYDN 11

RESULT 30
US-10-425-115-204064
; Sequence 204064, Application US/10425115
; Publication No. US2004021427A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```



APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 204064  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_117595C.1.pep  
US-10-425-115-204064

Query Match 40.2%; Score 43; DB 17; Length 109;  
Best Local Similarity 47.4%; Pred. No. 77;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19  
| : || | : || ||  
Db 21 RALYKVERGAAGGQEHKKG 39

## RESULT 31

US-10-767-701-50418  
; Sequence 50418, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 50418  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(154)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3480-006-Pl-K1-C5.pep  
US-10-767-701-50418

Query Match 40.2%; Score 43; DB 16; Length 154;  
Best Local Similarity 52.6%; Pred. No. 1.1e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

Qy 3 DYKFY--EDANGTRDHKKG 19  
| | | | | : || | : ||  
Db 103 DYDFYKESKANARRAKKRG 121

## RESULT 32

US-10-424-599-160723  
; Sequence 160723, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 160723  
; LENGTH: 274  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_116150C.1.pep  
US-10-424-599-160723

Query Match 40.2%; Score 43; DB 15; Length 274;  
Best Local Similarity 50.0%; Pred. No. 2e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKFYEDANGTRDHKKG 19  
| : || | : || ||  
Db 172 HKCQKEDGTDHKKG 187

## RESULT 33

US-10-767-701-39724  
; Sequence 39724, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 39724  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C28861\_1.pep  
US-10-767-701-39724

Query Match 40.2%; Score 43; DB 16; Length 337;  
Best Local Similarity 53.3%; Pred. No. 2.5e+02;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRD 15  
| | | | | : || | : ||  
Db 274 RSDQALYDGGNGTTD 288

## RESULT 34

US-10-437-963-128293  
; Sequence 128293, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 128293  
; LENGTH: 1299  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_30660C.1.pep  
US-10-437-963-128293

Query Match 40.2%; Score 43; DB 16; Length 1299;  
Best Local Similarity 50.0%; Pred. No. 9.7e+02;  
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 SDYKFYEDANGTRDHKKG 19  
:|||||:|||||  
Db 123 NDKKYEKGQGERVHNG 140

## RESULT 35

US-10-424-599-162795  
; Sequence 162795, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 162795  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_118021C.1.pep  
US-10-424-599-162795

Query Match 39.7%; Score 42.5; DB 15; Length 81;  
Best Local Similarity 47.4%; Pred. No. 68;  
Matches 9; Conservative 1; Mismatches 6; Indels 3; Gaps 1;

QY 4 YKF---YEDANGTRDHKKG 19  
|||:|:|:|:|:|:|  
Db 27 YMFVRVDYADQGRNEHKKG 45

## RESULT 36

US-10-424-599-285321  
; Sequence 285321, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 285321  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99674C.1.pep  
US-10-424-599-285321

Query Match 39.7%; Score 42.5; DB 15; Length 202;  
Best Local Similarity 60.0%; Pred. No. 1.7e+02;  
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 2 SDYKFYEDANGTRDH 16  
|||||:|

Db 31 SDYSF-SDANGDKGH 44

## RESULT 37

US-10-369-493-6568  
; Sequence 6568, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6568  
; LENGTH: 995  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_88090C.1.pep  
US-10-369-493-6568

Query Match 39.7%; Score 42.5; DB 14; Length 995;  
Best Local Similarity 42.1%; Pred. No. 8.8e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 RSDYKFYEDANGTRDHKKG 19  
:||||:|:|:|  
Db 257 RNDYRSQQD---SRDHSG 272

## RESULT 38

US-10-424-599-272502  
; Sequence 272502, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 272502  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_88090C.1.pep  
US-10-424-599-272502

Query Match 39.3%; Score 42; DB 15; Length 91;  
Best Local Similarity 42.1%; Pred. No. 92;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 RSDYKFYEDANGTRDHKKG 19  
:||||:|:|:|  
Db 30 KTDKKEENLDGERNHTKG 48

## RESULT 39

US-10-289-762-135  
; Sequence 135, Application US/10289762  
; Publication No. US20040006218A1  
; GENERAL INFORMATION:

```

; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 135
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-135

```

Query Match 39.3%; Score 42; DB 15; Length 189;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKFYEDANGTRDHHK 18  
| : | | : | | | |  
Db 58 DHKFDEETTTNRDHF 73

**RESULT 40**

```

US-10-282-122A-54860
; Sequence 54860, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54860
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-282-122A-54860

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:29:11 ; Search time 71.663 Seconds  
(without alignments)  
94.040 Million cell updates/sec

Title: US-09-719-379A-4  
Perfect score: 98  
Sequence: 1 RSDYKLYNKNSSSTLKDLGE 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PTCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                | Description       |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1          | 98    | 100.0       | 19     | US-10-203-942-4   | Sequence 4, Appli |
| 2          | 98    | 100.0       | 28     | US-10-203-942-8   | Sequence 8, Appli |
| 3          | 98    | 100.0       | 353    | US-10-467-421-21  | Sequence 21, Appl |
| 4          | 87    | 88.8        | 352    | US-10-336-840-37  | Sequence 37, Appl |
| 5          | 87    | 88.8        | 353    | US-10-203-942-9   | Sequence 9, Appli |
| 6          | 81.5  | 83.2        | 22     | US-10-203-942-2   | Sequence 2, Appli |
| 7          | 81.5  | 83.2        | 31     | US-10-203-942-6   | Sequence 6, Appli |
| 8          | 49    | 50.0        | 885    | US-09-867-852-2   | Sequence 2, Appli |
| 9          | 49    | 50.0        | 885    | US-10-613-472-2   | Sequence 2, Appli |
| 10         | 49    | 50.0        | 885    | US-10-613-765-2   | Sequence 2, Appli |
| 11         | 49    | 50.0        | 909    | US-09-867-852-142 | Sequence 142, App |
| 12         | 49    | 50.0        | 909    | US-10-613-472-142 | Sequence 142, App |
| 13         | 49    | 50.0        | 909    | US-10-613-765-142 | Sequence 142, App |

|    |    |      |      |    |                      |                    |
|----|----|------|------|----|----------------------|--------------------|
| 14 | 48 | 49.0 | 365  | 15 | US-10-282-122A-45498 | Sequence 45498, A  |
| 15 | 48 | 49.0 | 536  | 15 | US-10-282-122A-45804 | Sequence 45804, A  |
| 16 | 47 | 48.0 | 720  | 13 | US-10-008-355-9      | Sequence 9, Appli  |
| 17 | 46 | 46.9 | 648  | 15 | US-10-282-122A-52321 | Sequence 52321, A  |
| 18 | 46 | 46.9 | 1197 | 16 | US-10-437-963-167399 | Sequence 167399, A |
| 19 | 46 | 46.9 | 1829 | 14 | US-10-435-766-13     | Sequence 13, Appli |
| 20 | 45 | 45.9 | 624  | 14 | US-10-032-585-7797   | Sequence 7797, Ap  |
| 21 | 45 | 45.9 | 728  | 16 | US-10-437-963-132722 | Sequence 132722, A |
| 22 | 45 | 45.9 | 850  | 16 | US-10-437-963-132673 | Sequence 132673, A |
| 23 | 45 | 45.9 | 1127 | 16 | US-10-437-963-132719 | Sequence 132719, A |
| 24 | 45 | 45.9 | 1148 | 16 | US-10-437-963-132660 | Sequence 132660, A |
| 25 | 45 | 45.9 | 1234 | 16 | US-10-437-963-167391 | Sequence 167391, A |
| 26 | 45 | 45.9 | 1310 | 16 | US-10-437-963-132735 | Sequence 132735, A |
| 27 | 45 | 45.9 | 1335 | 16 | US-10-437-963-132777 | Sequence 132777, A |
| 28 | 45 | 45.9 | 1340 | 16 | US-10-437-963-132786 | Sequence 132786, A |
| 29 | 45 | 45.9 | 1370 | 16 | US-10-437-963-132773 | Sequence 132773, A |
| 30 | 45 | 45.9 | 1386 | 16 | US-10-437-963-132723 | Sequence 132723, A |
| 31 | 45 | 45.9 | 1391 | 16 | US-10-437-963-132727 | Sequence 132727, A |
| 32 | 45 | 45.9 | 1407 | 16 | US-10-437-963-132720 | Sequence 132720, A |
| 33 | 45 | 45.9 | 1479 | 16 | US-10-437-963-132825 | Sequence 132825, A |
| 34 | 45 | 45.9 | 1693 | 16 | US-10-437-963-132820 | Sequence 132820, A |
| 35 | 44 | 44.9 | 1058 | 14 | US-10-369-493-22153  | Sequence 22153, A  |
| 36 | 44 | 44.9 | 1216 | 14 | US-10-369-493-1505   | Sequence 1505, Ap  |
| 37 | 44 | 44.9 | 1287 | 16 | US-10-437-963-199679 | Sequence 199679, A |
| 38 | 44 | 44.9 | 1539 | 16 | US-10-437-963-112611 | Sequence 112611, A |
| 39 | 43 | 43.9 | 87   | 15 | US-10-424-599-195626 | Sequence 195626, A |
| 40 | 43 | 43.9 | 88   | 15 | US-10-424-599-276118 | Sequence 276118, A |
| 41 | 43 | 43.9 | 164  | 15 | US-10-335-977-9136   | Sequence 9136, Ap  |
| 42 | 43 | 43.9 | 263  | 15 | US-10-335-977-9137   | Sequence 9137, Ap  |
| 43 | 43 | 43.9 | 480  | 14 | US-10-369-493-2288   | Sequence 2288, Ap  |
| 44 | 43 | 43.9 | 1191 | 16 | US-10-437-963-169900 | Sequence 169900, A |
| 45 | 43 | 43.9 | 1285 | 16 | US-10-437-963-132816 | Sequence 132816, A |

ALIGNMENTS

RESULT 1  
US-10-203-942-4  
; Sequence 4, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOEL, PHILIPPE  
; APPLICANT: POOLMAN, JAN  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; FILE REFERENCE: B45210  
; CURRENT APPLICATION NUMBER: US/10/203,942  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-4

Query Match 100.0%; Score 98; DB 14; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 RSDYKLYNKNSSSTLKDLGE 19  
Db 1 RSDYKLYNKNSSSTLKDLGE 19

RESULT 2

```

US-10-203-942-8
; Sequence 8, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-8
Query Match      100.0%; Score 98; DB 14; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDIGE 19
| | | | | | | | | | | | | | | | | | | |
Db 3 RSDYKLYNKSSTLKDIGE 21

RESULT 3
US-10-467-421-21
; Sequence 21, Application US/10467421
; Publication No. US20040116665A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, Francois-Xavier Jacques
; APPLICANT: DENOEL, Philippe
; APPLICANT: NEYT, Cecile Anne
; APPLICANT: POOLMAN, Jan
; APPLICANT: THONNARD, Joelle
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45259
; CURRENT APPLICATION NUMBER: US/10/467,421
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01361
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103171.5
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(353)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-467-421-21
Query Match      100.0%; Score 98; DB 16; Length 353;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDIGE 19
| | | | | | | | | | | | | | | | | | | |
Db 132 RSDYKLYNKSSTLKDIGE 150

US-10-203-942-9
; Sequence 9, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-9
Query Match      88.8%; Score 87; DB 14; Length 353;
Best Local Similarity 89.5%; Pred. No. 3e-05;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDIGE 19
| | | | | | | | | | | | | | | | | | | |
Db 132 RSDYKLYNKSSTLKDIGE 150

RESULT 4
US-10-336-840-37
; Sequence 37, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-336-840-37
Query Match      88.8%; Score 87; DB 14; Length 352;
Best Local Similarity 89.5%; Pred. No. 3e-05;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDIGE 19
| | | | | | | | | | | | | | | | | | | |
Db 132 RSDYKLYNKSSTLKDIGE 150

RESULT 5
US-10-203-942-9
; Sequence 9, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT APPLICATION NUMBER: US/10/203,942
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-9
Query Match      88.8%; Score 87; DB 14; Length 353;
Best Local Similarity 89.5%; Pred. No. 3e-05;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDIGE 19
| | | | | | | | | | | | | | | | | | | |
Db 132 RSDYKLYNKSSTLKDIGE 150

RESULT 6
US-10-203-942-2
; Sequence 2, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:

```

APPLICANT: Brent, Andrew F.

|| : || : || : || :

```
Db      103 ADYKLCCKVSAILKSIGE 120

RESULT 10
; Sequence 2, Application US/10613765
; Publication No. US20040172673A1
; GENERAL INFORMATION:
; APPLICANT: AUSUBEL, FREDERICK M.
; APPLICANT: STASKAWICZ, BRIAN J.
; APPLICANT: KATAGIRI, FUMIAKI
; APPLICANT: KUNDEL, BARBARA N.
; APPLICANT: MINDRINOS, MICHAEL N.
; APPLICANT: YU, GUO-LIANG
; APPLICANT: BAKER, BARBARA
; APPLICANT: ELLIS, JEFFREY
; APPLICANT: SALMERON, JOHN
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
; TITLE OF INVENTION: DETECTION METHODS
; FILE REFERENCE: 00786/254005
; CURRENT APPLICATION NUMBER: US/10/613,765
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 09/867,852
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 09/301,085
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: US 08/310,912
; PRIOR FILING DATE: 1994-09-22
; PRIOR APPLICATION NUMBER: US 08/227,360
; PRIOR FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-613-765-2

Query Match      50.0%; Score 49; DB 16; Length 885;
Best Local Similarity 55.6%; Pred. No. 77;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy      2 SDYKLYNKSSTLKDGE 19
;||||| |:|:|:|
Db      103 ADYKLCCKVSAILKSIGE 120

RESULT 11
US-09-867-852-142
; Sequence 142, Application US/09867852
; Patent No. US20020147324A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
; TITLE OF INVENTION: DETECTION METHODS
; FILE REFERENCE: 00786/254002
; CURRENT APPLICATION NUMBER: US/09/867,852
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/301,085
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/310,912
; PRIOR FILING DATE: EARLIER FILING DATE: 1994-09-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/227,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 142
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-867-852-142

Query Match      50.0%; Score 49; DB 9; Length 909;
Best Local Similarity 55.6%; Pred. No. 80;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy      2 SDYKLYNKSSTLKDGE 19
;||||| |:|:|:|
Db      111 ADYKLCCKVSAILKSIGE 128

RESULT 12
US-10-613-472-142
; Sequence 142, Application US/10613472
; Publication No. US20040088756A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
; TITLE OF INVENTION: DETECTION METHODS
; FILE REFERENCE: 00786/254004
; CURRENT APPLICATION NUMBER: US/10/613,472
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 09/867,852
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 09/301,085
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: US 08/310,912
; PRIOR FILING DATE: 1994-09-22
; PRIOR APPLICATION NUMBER: US 08/227,360
; PRIOR FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 142
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-613-472-142

Query Match      50.0%; Score 49; DB 15; Length 909;
Best Local Similarity 55.6%; Pred. No. 80;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy      2 SDYKLYNKSSTLKDGE 19
;||||| |:|:~|:|
Db      111 ADYKLCCKVSAILKSIGE 128

RESULT 13
US-10-613-765-142
; Sequence 142, Application US/10613765
; Publication No. US20040172673A1
; GENERAL INFORMATION:
; APPLICANT: AUSUBEL, FREDERICK M.
; APPLICANT: STASKAWICZ, BRIAN J.
; APPLICANT: KATAGIRI, FUMIAKI
; APPLICANT: KUNDEL, BARBARA N.
; APPLICANT: MINDRINOS, MICHAEL N.
; APPLICANT: YU, GUO-LIANG
; APPLICANT: BAKER, BARBARA
; APPLICANT: ELLIS, JEFFREY
; APPLICANT: SALMERON, JOHN
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
```



;; TITLE OF INVENTION: DETECTION METHODS  
;; FILE REFERENCE: 00786/254005  
;; CURRENT APPLICATION NUMBER: US/10/613,765  
;; CURRENT FILING DATE: 2003-07-02  
;; PRIOR APPLICATION NUMBER: US 09/867,852  
;; PRIOR FILING DATE: 2001-05-29  
;; PRIOR APPLICATION NUMBER: US 09/301,085  
;; PRIOR FILING DATE: 1999-04-28  
;; PRIOR APPLICATION NUMBER: US 08/310,912  
;; PRIOR FILING DATE: 1994-09-22  
;; PRIOR APPLICATION NUMBER: US 08/227,360  
;; PRIOR FILING DATE: 1994-04-13  
;; NUMBER OF SEQ ID NOS: 214  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 142  
;; LENGTH: 909  
;; TYPE: PRT  
;; ORGANISM: Arabidopsis thaliana  
US-10-613-765-142

Query Match 50.0%; Score 49; DB 16; Length 909;  
Best Local Similarity 55.6%; Pred. No. 80;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLXDLGE 19  
Db 111 ADYKLCRKVSAILKSIGE 128

RESULT 14  
US-10-282-122A-45498  
; Sequence 45498, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 45498  
;; LENGTH: 365  
;; TYPE: PRT  
;; ORGANISM: Bacillus anthracis  
;; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (93)..(93)  
; OTHER INFORMATION: X=any amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (175)..(175)  
; OTHER INFORMATION: X=any amino acid  
US-10-282-122A-45498

Query Match 49.0%; Score 48; DB 15; Length 365;  
Best Local Similarity 71.4%; Pred. No. 41;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 YKLYNKSSTLKDL 17  
Db 14 YKLYNKPFDKLDL 27

RESULT 15  
US-10-282-122A-45804  
; Sequence 45804, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 45804  
;; LENGTH: 536  
;; TYPE: PRT  
;; ORGANISM: Bacillus anthracis  
;; FEATURE:  
; NAME/KEY: MISC FEATURE

; LOCATION: (369)..(369)  
 ; OTHER INFORMATION: X-any amino acid  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (371)..(371)  
 ; OTHER INFORMATION: X-any amino acid  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (406)..(406)  
 ; OTHER INFORMATION: X-any amino acid  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (420)..(420)  
 ; OTHER INFORMATION: X-any amino acid  
 US-10-282-122A-45804

Query Match 49.0%; Score 48; DB 15; Length 536;  
 Best Local Similarity 71.4%; Pred. No. 63;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 YKLYNKSSTLKDL 17  
 |||||  
 Db 14 YKLYNKPDKLKD 27

RESULT 16  
 US-10-008-355-9  
 ; Sequence 9, Application US/10008355  
 ; Publication No. US20020164759A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Travis, James  
 ; APPLICANT: Potempa, Jan S  
 ; APPLICANT: Banbula, Agnieszka  
 ; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
 ; FILE REFERENCE: 235.00440101  
 ; CURRENT APPLICATION NUMBER: US/10/008,355  
 ; CURRENT FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/246,827  
 ; PRIOR FILING DATE: 2000-11-08  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 9  
 ; LENGTH: 720  
 ; TYPE: PRT  
 ; ORGANISM: Porphyromonas gingivalis  
 US-10-008-355-9

Query Match 48.0%; Score 47; DB 13; Length 720;  
 Best Local Similarity 57.1%; Pred. No. 13e+02;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKD 16  
 :|||  
 Db 47 EYDLYNPNGTSLKD 60

RESULT 17  
 US-10-282-122A-52321  
 ; Sequence 52321, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 52321  
 ; LENGTH: 648  
 ; TYPE: PRT  
 ; ORGANISM: Clostridium botulinum  
 US-10-282-122A-52321

Query Match 46.9%; Score 46; DB 15; Length 648;  
 Best Local Similarity 47.4%; Pred. No. 1.6e+02;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19  
 |||:|  
 Db 422 RFDYLMYDKNDEFFKELCE 440

RESULT 18  
 US-10-437-963-167399  
 ; Sequence 167399, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 167399  
 ; LENGTH: 1197  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66013C.1.pep  
 US-10-437-963-167399

Query Match 46.9%; Score 46; DB 16; Length 1197;  
 Best Local Similarity 50.0%; Pred. No. 3.2e+02;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 724 RDKYKYNIEESHMED 739

US-10-435-766-13  
; Sequence 13, Application US/10435766  
; Publication No. US20030228616A1  
; GENERAL INFORMATION:  
; APPLICANT: Strategene  
; APPLICANT: Sorge, Joseph A  
; APPLICANT: Arzsi, Braham  
; APPLICANT: Hogrefe, Holly  
; APPLICANT: Hansen, Connie J  
; TITLE OF INVENTION: DNA Polymerase Mutants with Reverse Transcriptase Activity  
; FILE REFERENCE: 25436/1565C  
; CURRENT APPLICATION NUMBER: US/10/435,766  
; CURRENT FILING DATE: 2003-05-12  
; PRIOR APPLICATION NUMBER: US 10/223,650  
; PRIOR FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: US 09/896,923  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 09/698,341  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60/162,600  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: PCT/US00/29706  
; PRIOR FILING DATE: 2000-10-27  
; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 1829  
; TYPE: PRT  
; ORGANISM: Thermococcus strain TV  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1118)..(1118)  
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1123)..(1123)  
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
; US-10-435-766-13

Query Match 46.9%; Score 46; DB 14; Length 1829;  
Best Local Similarity 61.5%; Pred. No. 5.2e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KLYNKNSSTLKD 17  
| : : : : :  
Db 1161 KLYNENPNVLKDM 1173

US-10-032-585-7797  
; Sequence 7797, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7797  
; LENGTH: 624  
; TYPE: PRT  
; ORGANISM: Candida albicans

US-10-032-585-7797  
Query Match 45.9%; Score 45; DB 14; Length 624;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 18  
| : : : : :  
Db 69 RDKYKYNIEESHMED 86

RESULT 21  
US-10-437-963-132722  
; Sequence 132722, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132722  
; LENGTH: 728  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(728)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34663C.1.pep  
; US-10-437-963-132722

Query Match 45.9%; Score 45; DB 16; Length 728;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 187 RDKYKYNIEESHMED 202

US-10-437-963-132673  
; Sequence 132673, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132673  
; LENGTH: 850  
; TYPE: PRT

```
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(850)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34619C.1.pep
US-10-437-963-132673

Query Match      45.9%; Score 45; DB 16; Length 850;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16
   |||:|||||:|
Db 595 RDDYKVYNIESHMED 610

RESULT 23
US-10-437-963-132719
; Sequence 132719, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 132719
; LENGTH: 1127
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34660C.1.pep
US-10-437-963-132719

Query Match      45.9%; Score 45; DB 16; Length 1127;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16
   |||:|||||:|
Db 591 RDDYKVYNIESHMED 606

RESULT 24
US-10-437-963-132660
; Sequence 132660, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

```
; SEQ ID NO 132660
; LENGTH: 1148
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1148)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34607C.1.pep
US-10-437-963-132660

Query Match      45.9%; Score 45; DB 16; Length 1148;
Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16
   |||:|||||:|
Db 749 RDDYKVYNIESHMED 764

RESULT 25
US-10-437-963-167391
; Sequence 167391, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 167391
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66006C.1.pep
US-10-437-963-167391

Query Match      45.9%; Score 45; DB 16; Length 1234;
Best Local Similarity 50.0%; Pred. No. 4.8e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16
   |||:|||||:|
Db 698 RDDYKVYNIESHMED 713

RESULT 26
US-10-437-963-132735
; Sequence 132735, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
```

; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132735  
; LENGTH: 1310  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34675C.1.pep  
US-10-437-963-132735

Query Match 45.9%; Score 45; DB 16; Length 1310;  
Best Local Similarity 50.0%; Pred. No. 5.1e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 840 RDDYKVYNIIESHMD 855

RESULT 27  
US-10-437-963-132777  
; Sequence 132777, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132777  
; LENGTH: 1335  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34712C.1.pep  
US-10-437-963-132777

Query Match 45.9%; Score 45; DB 16; Length 1335;  
Best Local Similarity 50.0%; Pred. No. 5.2e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 799 RDDYKVYNIIESHMD 814

RESULT 28  
US-10-437-963-132786  
; Sequence 132786, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132786  
; LENGTH: 1340  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1340)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34720C.1.pep  
US-10-437-963-132786

Query Match 45.9%; Score 45; DB 16; Length 1340;  
Best Local Similarity 50.0%; Pred. No. 5.2e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 804 RDDYKVYNIIESHMD 819

RESULT 29  
US-10-437-963-132773  
; Sequence 132773, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132773  
; LENGTH: 1370  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34709C.1.pep  
US-10-437-963-132773

Query Match 45.9%; Score 45; DB 16; Length 1370;  
Best Local Similarity 50.0%; Pred. No. 5.4e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKD 16  
| : : : : :  
Db 834 RDDYKVYNIIESHMD 849

RESULT 30  
US-10-437-963-132723  
; Sequence 132723, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132723  
; LENGTH: 1386  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34664C.1.pep  
US-10-437-963-132723

Query Match 45.9%; Score 45; DB 16; Length 1386;  
Best Local Similarity 50.0%; Pred. No. 5.4e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
| |||: || | : ||  
Db 850 RDDYKVYNIESHMED 865

RESULT 31  
US-10-437-963-132727  
; Sequence 132727, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132727  
; LENGTH: 1391  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34668C.1.pep  
US-10-437-963-132727

Query Match 45.9%; Score 45; DB 16; Length 1391;  
Best Local Similarity 50.0%; Pred. No. 5.4e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
| |||: || | : ||  
Db 855 RDDYKVYNIESHMED 870

RESULT 32  
US-10-437-963-132720  
; Sequence 132720, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132720  
; LENGTH: 1407  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34661C.1.pep  
US-10-437-963-132720

Query Match 45.9%; Score 45; DB 16; Length 1407;  
Best Local Similarity 50.0%; Pred. No. 5.5e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
| |||: || | : ||  
Db 871 RDDYKVYNIESHMED 886

RESULT 33  
US-10-437-963-132825  
; Sequence 132825, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132825  
; LENGTH: 1479  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34756C.1.pep  
US-10-437-963-132825

Query Match 45.9%; Score 45; DB 16; Length 1479;  
Best Local Similarity 50.0%; Pred. No. 5.8e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
| |||: || | : ||  
Db 943 RDDYKVYNIESHMED 958

RESULT 34  
US-10-437-963-132820  
; Sequence 132820, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 132820  
; LENGTH: 1693  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_34751C.1.pep  
US-10-437-963-132820

Query Match 45.9%; Score 45; DB 16; Length 1693;  
Best Local Similarity 50.0%; Pred. No. 6.8e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
Db 833 RDDYKVYNIESHMED 848

RESULT 35  
US-10-369-493-22153  
; Sequence 22153, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22153  
; LENGTH: 1058  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22153

Query Match 44.9%; Score 44; DB 14; Length 1058;  
Best Local Similarity 60.0%; Pred. No. 5.8e+02;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKDL 17  
Db 72 DYKAYKPNLSLNDL 86

RESULT 36  
US-10-369-493-1505  
; Sequence 1505, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1505  
; LENGTH: 1216

; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1505

Query Match 44.9%; Score 44; DB 14; Length 1216;  
Best Local Similarity 57.1%; Pred. No. 6.7e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDLG 18  
Db 284 RMYNRYSSILKNLG 297

RESULT 37  
US-10-437-963-199679  
; Sequence 199679, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 199679  
; LENGTH: 1287  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_95222C.1.pep  
US-10-437-963-199679

Query Match 44.9%; Score 44; DB 16; Length 1287;  
Best Local Similarity 50.0%; Pred. No. 7.2e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTL 14  
Db 1114 RDDYKYVDQNEKNL 1127

RESULT 38  
US-10-437-963-112611  
; Sequence 112611, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 112611  
; LENGTH: 1539  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1647C.1.pep  
US-10-437-963-112611

Query Match 44.9%; Score 44; DB 16; Length 1539;  
Best Local Similarity 37.5%; Pred. No. 8.8e+02;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 4 YKLYNKNSSTLKDLGE 19  
Db 922 YRVNKNKGTVETAD 937

RESULT 39

US-10-424-599-195626  
; Sequence 195626, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 195626  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_18677C.1.pep  
US-10-424-599-195626

Query Match 43.9%; Score 43; DB 15; Length 87;  
Best Local Similarity 57.1%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLK 15  
Db 54 SNHLYNKNSPQLR 67

RESULT 40

US-10-424-599-276118  
; Sequence 276118, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 276118  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(88)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_91354C.1.pep  
US-10-424-599-276118

Query Match 43.9%; Score 43; DB 15; Length 88;  
Best Local Similarity 53.3%; Pred. No. 51;

Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
Qy 3 DYKLYNKNSSTLKDL 17  
Db 5 EYKIWKNSXFLYDL 19  
Search completed: November 24, 2004, 10:00:15  
Job time : 72.663 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:04 ; Search time 87.3587 Seconds  
(without alignments)  
125.141 Million cell updates/sec

Title: US-09-719-379A-5  
Perfect score: 107  
Sequence: 1 RSDYKFYEDANGTRDHKKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 02.\*

- 1: uniprot\_sprot.\*
- 2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description        |
|------------|-------|---------------|--------|--------------|--------------------|
| 1          | 107   | 100.0         | 353    | 1 OM52_HAEIN | P38368 haemophilus |
| 2          | 107   | 100.0         | 359    | 1 OM53_HAEIN | P45996 haemophilus |
| 3          | 56.5  | 52.8          | 360    | 2 Q86254     | Q86254 haemophilus |
| 4          | 49    | 45.8          | 128    | 2 Q8ESU7     | Q8ESU7 oceanobacil |
| 5          | 49    | 45.8          | 294    | 2 Q9NWN0     | Q9NWN0 homo sapien |
| 6          | 49    | 45.8          | 305    | 2 Q9H278     | Q9H278 homo sapien |
| 7          | 49    | 45.8          | 470    | 2 Q9NWK9     | Q9NWK9 homo sapien |
| 8          | 49    | 45.8          | 470    | 2 Q9H3X3     | Q9H3X3 homo sapien |
| 9          | 49    | 45.8          | 536    | 2 Q72BV4     | Q72BV4 desulfovibr |
| 10         | 49    | 45.8          | 536    | 2 AAS96008   | AAS96008 desulfovi |
| 11         | 47.5  | 44.4          | 366    | 2 Q7AE32     | Q7AE32 escherichia |
| 12         | 47.5  | 44.4          | 366    | 2 Q8XAS0     | Q8XAS0 escherichia |
| 13         | 47.5  | 44.4          | 997    | 2 Q9GYL4     | Q9GYL4 caenorhabdi |
| 14         | 47    | 43.9          | 304    | 2 Q9LNL9     | Q9LNL9 arabidopsis |
| 15         | 47    | 43.9          | 343    | 2 Q8GX84     | Q8GX84 arabidopsis |
| 16         | 47    | 43.9          | 378    | 2 Q8L789     | Q8L789 arabidopsis |
| 17         | 47    | 43.9          | 378    | 2 Q9FNG6     | Q9FNG6 arabidopsis |
| 18         | 47    | 43.9          | 751    | 2 Q8RDL1     | Q8RDL1 thermococ   |
| 19         | 47    | 43.9          | 868    | 1 PRTT_PORGI | P3158 porphyronon  |
| 20         | 47    | 43.9          | 886    | 2 Q53481     | Q53481 porphyronon |
| 21         | 47    | 43.9          | 1211   | 2 Q76YV0     | Q76YV0 bacterioph  |
| 22         | 47    | 43.9          | 1211   | 2 AAG17796   | AAG17796 bacteriop |
| 23         | 47    | 43.9          | 1467   | 2 Q6CVA3     | Q6CVA3 kluyveromyc |
| 24         | 47    | 43.9          | 1562   | 1 SPAP_STRMU | P33504 streptococ  |
| 25         | 47    | 43.9          | 1565   | 1 PAC_STRMU  | P11657 streptococ  |
| 26         | 47    | 43.9          | 1565   | 2 BAC54564   | BAC54564 streptoc  |
| 27         | 46.5  | 43.5          | 698    | 2 Q6HA27     | Q6HA27 trypanosoma |
| 28         | 46    | 43.0          | 180    | 2 Q6NMJ2     | Q6NMJ2 brachydanio |
| 29         | 46    | 43.0          | 180    | 2 AAB67570   | AAB67570 brachydan |
| 30         | 46    | 43.0          | 908    | 2 Q9LVX1     | Q9LVX1 arabidopsis |
| 31         | 46    | 43.0          | 908    | 2 Q8W4G4     | Q8W4G4 arabidopsis |

|    |      |      |      |   |            |                    |
|----|------|------|------|---|------------|--------------------|
| 32 | 46   | 43.0 | 1019 | 2 | Q74LR8     | Q74LR8 lactobacill |
| 33 | 46   | 43.0 | 1019 | 2 | AAS08097   | AAS08097 lactobaci |
| 34 | 46   | 43.0 | 1090 | 2 | Q928J2     | Q928J2 listeria in |
| 35 | 46   | 43.0 | 1091 | 2 | Q8Y4J2     | Q8Y4J2 listeria mo |
| 36 | 46   | 43.0 | 1091 | 2 | Q71WY0     | Q71WY0 listeria mo |
| 37 | 46   | 43.0 | 1091 | 2 | AAT05185   | AAT05185 listeria  |
| 38 | 45.5 | 42.5 | 345  | 1 | YG3V_YEAST | Y53292 saccharomyc |
| 39 | 45.5 | 42.5 | 381  | 2 | Q8KAK9     | Q8KAK9 chlorobium  |
| 40 | 45.5 | 42.5 | 890  | 2 | Q8B488     | Q8B488 iguape viru |
| 41 | 45   | 42.1 | 277  | 2 | Q6NYV4     | Q6NYV4 brachydanio |
| 42 | 45   | 42.1 | 277  | 2 | AAB66448   | AAB66448 brachydan |
| 43 | 45   | 42.1 | 312  | 2 | Q6SPF9     | Q6SPF9 uncultured  |
| 44 | 45   | 42.1 | 312  | 2 | AAR38163   | AAR38163 unculture |
| 45 | 45   | 42.1 | 339  | 2 | Q98S79     | Q98S79 guillardia  |

ALIGNMENTS

RESULT 1

| ID | OM52_HAEIN   | STANDARD; | PRT; | 353 AA. |
|----|--|-----------|------|---------|
| AC | P38368;  |           |      |         |
| DT | 01-OCT-1994 (Rel. 30, Created)   |           |      |         |
| DT | 01-OCT-1994 (Rel. 30, Last sequence update)                                |           |      |         |
| DT | 05-JUL-2004 (Rel. 44, Last annotation update)                              |           |      |         |
| DE | Outer membrane protein P5 precursor (OMP P5).                              |           |      |         |
| GN | Name=ompA; Synonyms=ompP5;   |           |      |         |
| OS | Haemophilus influenzae.  |           |      |         |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;             |           |      |         |
| OC | Pasteurellaceae; Haemophilus.  |           |      |         |
| OX | NCBI_TaxID=727;  |           |      |         |
| RN | [1]  |           |      |         |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 22-34.                                 |           |      |         |
| RC | STRAIN=1613 / Serotype B;  |           |      |         |
| RX | MEDLINE=93366472; PubMed=8359929;  |           |      |         |
| RA | Munson R.S. Jr.; Grass S.; West R.;  |           |      |         |
| RT | "Molecular cloning and sequence of the gene for outer membrane protein     |           |      |         |
| RL | P5 of Haemophilus influenzae.";  |           |      |         |
| RL | Infect. Immun. 61:4017-4020(1993).   |           |      |         |
| CC | -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.       |           |      |         |
| CC | -1- SIMILARITY: Belongs to the ompA family.                                |           |      |         |
| CC |  |           |      |         |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |           |      |         |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -    |           |      |         |
| CC | the European Bioinformatics Institute. There are no restrictions on its    |           |      |         |
| CC | use by non-profit institutions as long as its content is in no way         |           |      |         |
| CC | modified and this statement is not removed. Usage by and for commercial    |           |      |         |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ |           |      |         |
| CC | or send an email to license@isb-sib.ch).                                   |           |      |         |
| CC |  |           |      |         |
| DR | EMBL; L20309; AAA03346.1; -.   |           |      |         |
| DR | HSP; P02934; 1QJP.   |           |      |         |
| DR | InterPro; IPR006664; Bac OmpA.   |           |      |         |
| DR | InterPro; IPR002368; OmpA.   |           |      |         |
| DR | InterPro; IPR006665; OmpA/MotB.  |           |      |         |
| DR | InterPro; IPR006690; OMPA LIKE.  |           |      |         |
| DR | InterPro; IPR00498; OmpA_tmem.   |           |      |         |
| DR | Pfam; PF00691; OmpA; 1.  |           |      |         |
| DR | Pfam; PF01389; OmpA membrane; 1.   |           |      |         |
| DR | PRINTS; PR01021; OMPADOMAIN.   |           |      |         |
| DR | ProDom; PD000930; OmpA/MotB; 1.  |           |      |         |
| DR | PROSITE; PS01068; OMPA; 1.   |           |      |         |
| KW | Direct protein sequencing; Outer membrane; Porin; Signal;                  |           |      |         |
| KW | Transmembrane.   |           |      |         |
| FT | SIGNAL 1 21  |           |      |         |
| FT | CHAIN 22 353   |           |      |         |
| FT | DISULFID 326 338   |           |      |         |
| FT | DOMAIN 272 316   |           |      |         |
| SQ | SEQUENCE 353 AA; 37594 MW; E58A659E786D0F7 CRC64;                          |           |      |         |

Query Match 100.0%; Score 107; DB 1; Length 353;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;

```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19
    |||||
Db 132 RSDYKFYEDANGTRDHKKG 150

RESULT 2
OM53 HAEIN
ID _OM53 HAEIN STANDARD; PRT; 359 AA.
AC P45956;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Outer membrane protein P5 precursor (OMP P5) (Fimbrin).
GN Name=OmpA; Synonyms=ompP5;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NTHI 1128;
RA MEDLINE=9422575; PubMed=7909539;
RA Sirakova T., Kolattukudy P.E., Murwin D., Billy J., Leake E., Lim D.,
RA Denaria T., Bakaletz L.;
RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
RT pathogenesis of and protection against otitis media and relatedness of
RT the fimbrin subunit to outer membrane protein A.";
RL Infect. Immun. 62:2002-2020(1994).
CC -!- FUNCTION: Acts as a fimbriae subunit.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: Belongs to the ompA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi-sib.ch/announce/
CC or send an email to license@ebi-sib.ch).
CC -----
DR EMBL; L08448; AAA24959.1; -.
DR HSSP; P02934; 1BXW.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR002368; OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OMPA_LIKE.
DR InterPro; IPR000498; OmpA_tmem.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01389; OmpA_membrane; 1.
DR PRODOM; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OMPA; FALSE NEG.
KW Direct protein sequencing; Fimbria; Outer membrane; Porin; Signal;
KW Transmembrane.
FT SIGNAL 1 21
FT CHAIN 22 359 Outer membrane protein P5.
FT DISULFID 332 344 By similarity.
FT DOMAIN 278 322 OmpA-like.
SQ SEQUENCE 359 AA; 38340 MW; 576B1C59B4818C37 CRC64;

Query Match 100.0%; Score 107; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHKKG 19
    |||||
Db 138 RSDYKFYEDANGTRDHKKG 156

RESULT 3
O86254
ID _O86254 PRELIMINARY; PRT; 360 AA.
AC O86254;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Outer membrane protein (Fragment).
GN Name=omp;
OS Haemophilus sp.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=740;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16N;
RA MEDLINE=99081716; PubMed=9864189;
RA Goussset N., Rosenau A., Sizaret P.Y., Quentin R.;
RT "Nucleotide sequences of genes coding for fimbrial proteins in a
RT cryptic genospecies of Haemophilus spp. isolated from neonatal and
RT genital tract infections.";
RL Infect. Immun. 67:8-15(1999).
CC -!- SIMILARITY: Belongs to the ompA family.
CC EMBL; AJ007317; CAA07454.1; -.
CC GO; GO:0016021; C:integral to membrane; IEA.
CC GO; GO:0009279; C:outer membrane (sensu Gram-negative Bacteria); IEA.
CC GO; GO:0005198; F:structural molecule activity; IEA.
CC InterPro; IPR006664; Bac_OmpA.
CC InterPro; IPR002368; OmpA.
CC InterPro; IPR006665; OmpA/MotB.
CC InterPro; IPR006690; OMPA_LIKE.
CC InterPro; IPR000498; OmpA_tmem.
CC Pfam; PF00691; OmpA; 1.
CC PRINTS; PR01389; OmpA_membrane; 1.
CC PRINTS; PR01021; OMPADOMAIN.
CC PRODOM; PD000930; OmpA/MotB; 1.
CC PROSITE; PS01068; OMPA; 1.
FT NON TER 360 360
SQ SEQUENCE 360 AA; 38415 MW; A3209155051CDD69 CRC64;

Query Match 52.8%; Score 56.5; DB 2; Length 360;
Best Local Similarity 85.7%; Pred. No. 1.6;
Matches 12; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 RSDYKFYEDANGTR 14
    |||||
Db 140 RSDYKFY-DANGAR 152

RESULT 4
Q8ESU7
ID Q8ESU7 PRELIMINARY; PRT; 128 AA.
AC Q8ESU7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ectoine synthase.
GN OrderedLocuNames=OB0519;
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831;
RC MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004594; BAC12475.1; -.
DR InterPro; IPR010462; Ectoine_synth.
DR InterPro; IPR011051; RmlC_like_cupin.
DR Pfam; PF06339; Ectoine_synth; 1.

```

208 SDRYFLEDVARTADH 222

RESULT 7  
Q9NWK9 PRELIMINARY; PRT; 470 AA.

ID Q9NWK9  
AC Q9NWK9  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein FLJ20760 (Hypothetical protein FLJ20729).  
GN Names:FLJ20729;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA - Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isoigai T., Sugano S.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Testis;  
RC MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalka U., Smalusz D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Testis;  
RC Strausberg R.;  
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AK000767; BAA91371.1; -;  
DR EMBL; BC026236; AAB26236.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 470 AA; 53918 MW; 34A50A1C62B4F4E8 CRC64;  
Query Match 45.8%; Score 49; DB 2; Length 470;  
Best Local Similarity 60.0%; Pred. No. 33;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 SDYKPYEDANGTRDH 16  
|||||  
DB 276 SDRYFLEDVARTADH 290

RESULT 8  
Q9H3X3 PRELIMINARY; PRT; 470 AA.

ID Q9H3X3  
AC Q9H3X3  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein DKFZp434M1929.

|                       |  |
|-----------------------|--|
| GN                    | Name=DKFZp434M1929;  |
| OS                    | Homo sapiens (Human).  |
| OC                    | Eukaryota; Metazoa;  |
| OC                    | Cnidaria; Vertebrata; Euteleostomi;                                      |
| OC                    | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.               |
| OX                    | NCBI_TaxID=9606;   |
| RN                    | [1]  |
| RP                    | SEQUENCE FROM N.A.   |
| RC                    | TISSUE=Testis;   |
| RA                    | Koehrer K., Beyer A., Mewes H.W., Weil B., Wiemann S.;                   |
| RL                    | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.                  |
| DR                    | EMBL; AL442074; CAC09440.1; -.   |
| KW                    | Hypothetical protein.  |
| SQ                    | SEQUENCE 470 AA; 53846 MW; 9887ALA99C05C9E2 CRC64;                       |
| <br>                  |  |
| Query Match           | 45.8%; Score 49; DB 2; Length 470;                                       |
| Best Local Similarity | 60.0%; Pred.No. 33;  |
| Matches               | 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;                       |
| <br>                  |  |
| Qy                    | 2 SDYKFYEDANGTRDH 16<br>   :   |
| Dd                    | 276 SDYRFLEDVARTADH 290<br>   :  |
| <br>                  |  |
| RESULT 9              |  |
| Q72BV4                |  |
| ID                    | Q72BV4 PRELIMINARY; PRT; 536 AA.   |
| AC                    | Q72BV4;  |
| DT                    | 05-JUL-2004 (TrEMBLrel. 27, Created)                                     |
| DT                    | 05-JUN-2004 (TrEMBLrel. 27, Last sequence update)                        |
| DT                    | 03-JUL-2004 (TrEMBLrel. 27, Last annotation update)                      |
| DE                    | Metallo-beta-lactamase family protein.                                   |
| OR                    | OrderedLocusNames=DVU1530;   |
| OS                    | Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303). |
| OC                    | Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;       |
| OC                    | Desulfovibrionaceae; Desulfovibrio.                                      |
| OX                    | NCBI_TaxID=882;  |
| RN                    | [1]  |
| RP                    | SEQUENCE FROM N.A.   |
| RX                    | PubMed=15077118; DOI=10.1038/nbt959;                                     |
| RA                    | Heidelberg J.F., Seashadi R., Haveman S.A., Henne C.L., Paulsen I.T.,    |
| RA                    | Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,           |
| RA                    | Laughey S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,           |
| RA                    | Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,          |
| RA                    | Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,               |
| RA                    | Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J.J., Uterback T.R.,   |
| RA                    | Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;                     |
| RT                    | "The genome sequence of the anaerobic, sulfate-reducing bacterium        |
| RT                    | Desulfovibrio vulgaris Hildenborough.";                                  |
| RL                    | Nat. Biotechnol. 22:554-559(2004).                                       |
| DR                    | EMBL; AE017314; AAS96008.1; -.   |
| DR                    | TIGR; DVU1530; -.  |
| DR                    | InterPro; IPR001279; Blactmase-like.                                     |
| DR                    | InterPro; IPR011108; RMML.   |
| DR                    | Pfam; PF00753; Lactamase_B_1.  |
| DR                    | Pfam; PF07521; RMML; 1.  |
| KW                    | Complete proteome.   |
| SQ                    | SEQUENCE 536 AA; 59893 MW; C6D54A537BBCC286 CRC64;                       |
| <br>                  |  |
| Query Match           | 45.8%; Score 49; DB 2; Length 536;                                       |
| Best Local Similarity | 56.2%; Pred.No. 38;  |
| Matches               | 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;                       |
| <br>                  |  |
| Qy                    | 2 SDYKFYEDANGTRDH 17<br>:  |
| Dd                    | 206 ADYLFLSTYGDRDHK 221<br>:   |
| <br>                  |  |
| RESULT 10             |  |
| AAS96008              |  |
| ID                    | AAS96008 PRELIMINARY; PRT; 536 AA.                                       |
| AC                    | AAS96008;  |
| DT                    | 27-APR-2004 (TrEMBLrel. 27, Created)                                     |

```

DR ProDom; PD000808; OMP 2; 1.
KW PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Membrane; Outer membrane; Porin; Transmembrane.
SQ SEQUENCE 366 AA; 41025 MW; 9B16735B9BA9F25 CRC64;

Query Match
Best Local Similarity 44.4%; Score 47.5; DB 2; Length 366;
Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 3 DYKFYED-ANGTRDHK 17
Db 80 DFKGYNDEANGSRDKNK 95

RESULT 12
Q8XASO Q8XASO PRELIMINARY; PRT; 366 AA.
AC Q8XASO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Putative outer membrane porin protein.
GN OrderedLocustNames=z2239;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11208551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Godbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the Gram-negative porin family.
DR EMBL; AE005358; AAG56297.1; -.
DR PIR; D90888; D90888.
DR PIR; E85729; E85729.
DR HSSP; P02931; 1GFN.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005741; C: mitochondrial outer membrane; IEA.
DR GO; GO:0015288; F: porin activity; IEA.
DR GO; GO:0005215; F: transporter activity; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR003229; OMP 2.
DR InterPro; IPR001897; Porin bac.
DR InterPro; IPR001702; Porin Gram-ve.
DR Pfam; PF00267; Porin 1; 1.
DR PRINTS; PR00183; ECOLIPORIN.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP 2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Complete proteome; Membrane; Outer membrane; Porin; Transmembrane.
SQ SEQUENCE 366 AA; 41025 MW; 9B16735B9BA9F25 CRC64;

Query Match
Best Local Similarity 44.4%; Score 47.5; DB 2; Length 366;
Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 3 DYKFYED-ANGTRDHK 17
Db 80 DFKGYNDEANGSRDKNK 95

RESULT 13

```

```

Q9GYL4 Q9GYL4 PRELIMINARY; PRT; 997 AA.
AC Q9GYL4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hypothetical protein R04E5.8.
DE ORFNames=R04E5.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Miller N.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41538; AAG00010.1; -.
DR PIR; T28872; T28872.
DR WormPep; R04E5.8; CE04800.
KW Hypothetical protein.
SQ SEQUENCE 997 AA; 111954 MW; F1620378EF0D9DB6 CRC64;

Query Match
Best Local Similarity 44.4%; Score 47.5; DB 2; Length 997;
Matches 10; Conservative 2; Mismatches 1; Indels 3; Gaps 1;

Qy 2 SDYKFY---EDANGTR 14
Db 289 SGYRFYNGEDANGTK 304

RESULT 14
Q9LNI9 Q9LNI9 PRELIMINARY; PRT; 304 AA.
AC Q9LNI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Putative zinc finger protein.
GN Name=F6F3.14;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chiu J., Choi E., Gonzales A.,
RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaynsberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

```

CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 DR EMBL; AC023628; AAF97335.1; -.  
 DR PIR; G86143; G86143.  
 DR HSSP; P38398; 1JW7.  
 DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0016567; P:protein ubiquitination; IEA.  
 DR InterPro; IPR000571; Znf CCH.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00097; ZF-C3HC4; 1.  
 DR Pfam; PF00642; ZF-CCCH; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00356; Znf C3H1; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 304 AA; 34215 MW; 7C107FA164251D27 CRC64;

Query Match 43.9%; Score 47; DB 2; Length 304;  
 Best Local Similarity 50.0%; Pred. No. 43;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKPYEDANGTRDHKKG 19

Db 151 SDEKLYKGHYGTHKAG 168

RESULT 15

ID Q8GX84 PRELIMINARY; PRT; 343 AA.  
 AC Q8GX84;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein Atg01350/F6F3\_27.  
 GN Name=Atg01350/F6F3\_27;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 DR EMBL; AK118376; BAC42988.1; -.  
 DR HSSP; P15919; 1RMD.  
 DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0016567; P:protein ubiquitination; IEA.  
 DR InterPro; IPR000571; Znf CCH.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00097; ZF-C3HC4; 1.  
 DR Pfam; PF00642; ZF-CCCH; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00356; Znf C3H1; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Hypothetical protein; Metal-binding; Zinc; zinc-finger.  
 SQ SEQUENCE 343 AA; 38719 MW; E333D70369C31A83 CRC64;

Query Match 43.9%; Score 47; DB 2; Length 343;  
 Best Local Similarity 50.0%; Pred. No. 50;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKPYEDANGTRDHKKG 19

Db 151 SDEKLYKGHYGTHKAG 168

RESULT 16

Q8L789 PRELIMINARY; PRT; 378 AA.  
 AC Q8L789;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein At5g06420.  
 GN Name=At5g06420;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,  
 RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,  
 RA Deng J.M., Hayashizaki Y., Huan V.W., Lee J.M., Ishida J., Kamiya A.,  
 RA Kawai J., Kim C.J., Narusaka M., Quach H.L., Sakurai T., Satou M.,  
 RA Seki M., Shinn P., Tang C.C., Toroumi M., Wallender E.K., Wong C.,  
 RA Wu H.C., Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J.,  
 RA Theologis A., Davis R.W.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 DR EMBL; AY136406; AAM97072.1; -.  
 DR EMBL; BT000229; AAN15548.1; -.  
 DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0016567; P:protein ubiquitination; IEA.  
 DR InterPro; IPR000571; Znf CCH.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00097; ZF-C3HC4; 1.  
 DR Pfam; PF00642; ZF-CCCH; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00356; Znf C3H1; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Hypothetical protein; Metal-binding; Zinc; zinc-finger.  
 SQ SEQUENCE 378 AA; 42518 MW; 142E4A6534BECA4D CRC64;

Query Match 43.9%; Score 47; Length 378;  
 Best Local Similarity 50.0%; Pred. No. 55;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKPYEDANGTRDHKKG 19

Db 186 SDEKLYKGHYGTHKAG 203

RESULT 17

Q9FNG6 PRELIMINARY; PRT; 378 AA.  
 AC Q9FNG6;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 05-JUN-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Similarity to zinc finger protein (Hypothetical protein).

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OK NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98069011; PubMed=9405937;  
 RA Kobani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.  
 RT Sequence features of the regions of 1,044,062 bp covered by thirteen  
 RT physically assigned P1 clones.";  
 RL DNA Res. 4:291-300(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22088475; PubMed=12093376;  
 RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RT annotation.";  
 RL Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.  
 DR EMBL; AB006700; BAB08964.1; -;  
 DR EMBL; AY087435; AAM67329.1; -;  
 DR GO; GO:000151; C:ubiquitin ligase complex; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0016567; P:protein ubiquitination; IEA.  
 DR InterPro; IPR000571; Znf\_CCH.  
 DR InterPro; IPR001841; Znf\_Ring.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR Pfam; PF00642; zf-CCCH; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00356; Znf\_C3H1; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; UNKNOWN\_1.  
 DR PROSITE; PS00089; ZF\_RING\_2; 1.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 378 AA; 42460 MW; 173D71BB88A3FE2D CRC64;  
 Query Match 43.9%; Score 47; DB 2; Length 378;  
 Best Local Similarity 50.0%; Pred. No. 55;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
 Qy 2 SDYKPYEDANGTRDHKKG 19  
 Db 186 SDEKLYKGIHYTDHKG 203  
 ID Q8RDL1 PRELIMINARY; PRT; 751 AA.  
 AC Q8RDL1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE Alpha-glucosidases, family 31 of glycosyl hydrolases.  
 GN OrderedLocustNames=TT00006;  
 OS Thermoanaerobacter tengcongensis.  
 CC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
 CC Thermoanaerobacteriaceae; Thermoanaerobacter.  
 OK NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MB4;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,

RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of the T. tengcongensis genome.";  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL; AE012974; AAM23323.1; -;  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 KW Complete proteome; Hydrolase.  
 SQ SEQUENCE 751 AA; 88568 MW; 56AAFC74ED1ACFC9 CRC64;  
 Query Match 43.9%; Score 47; DB 2; Length 751;  
 Best Local Similarity 52.6%; Pred. No. 1.2e+02;  
 Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 Qy 1 RSDYKPYEDANGTRDHKKG 19  
 Db 671 RGRYHYEDDKTDFYKKG 689  
 ID -PRTT\_PORGI STANDARD; PRT; 868 AA.  
 AC P43158;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Thiol protease/hemagglutinin prtT precursor (EC 3.4.22.-).  
 GN Name=prtT;  
 OS Porphyromonas gingivalis (Bacteroides gingivalis).  
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 OC Porphyromonadaceae; Porphyromonas.  
 OK NCBI\_TaxID=837;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 53977;  
 RX MEDLINE=93114862; PubMed=8093357;  
 RA Ootogoto J.-I., Kuramitsu H.K.;  
 RT "Isolation and characterization of the Porphyromonas gingivalis prtT  
 RT gene, coding for protease activity";  
 RL Infect. Immun. 61:117-123(1993).  
 RN [2]  
 RP REVISIONS, SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 53977;  
 RX MEDLINE=95105001; PubMed=7806362;  
 RA Madden T.E., Clark V.L., Kuramitsu H.K.;  
 RT "Revised sequence of the Porphyromonas gingivalis prtT cysteine  
 RT protease/hemagglutinin gene: homology with streptococcal pyrogenic  
 RT exotoxin B/streptococcal proteinase.";  
 RL Infect. Immun. 63:238-247(1995).  
 CC -1- FUNCTION: Appears to be specific for arginine-containing peptide  
 CC bonds. Possesses hemagglutinin activity. C10.  
 CC -1- SIMILARITY: Belongs to peptidase family C10.  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-17 is the initiator.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/  
 CC or send an email to license@ebi.ac.uk).  
 CC -----  
 CC EMBL; M83096; -; NOT\_ANNOTATED\_CDS.  
 DR MEROPS; C10.002; -;  
 DR InterPro; IPR000200; Peptidase\_C10.  
 DR Pfam; PF01640; Peptidase\_C10; 1.  
 DR PRINTS; PR00797; STREPTOPAIN.  
 DR ProDom; PD004169; Peptidase\_C10; 1.  
 KW Hemagglutinin; Hydrolase; Signal; Thiol protease.

Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

Query Match 43.98; Score 47; DB 2; Length 1211;  
Best Local Similarity 39.18; Pred. No. 1.9e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 4;

Query 1 RSDYKFYEDANG----TRDHKKG 19  
DB 773 KNGFKFFNAENGKSISVRDHKKG 795

RESULT 22

AAQ17796  
ID AC AAQ17796 PRELIMINARY; PRT; 1211 AA.  
AC AAQ17796;  
DT 02-MAR-2004 (TRENBLrel. 27, Created)  
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN AEHIORFI31C.  
OS Bacteriophage Aehl.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC T4-like viruses.  
ON NCBI\_TaxID=227470;  
OX [1]  
RP SEQUENCE FROM N.A.  
RA Petrov V., Nolan J., Bertrand C., Letarov A.V., Kriech H.M.,  
RA Karam J.D.;  
RT "Aeromonas hydrophila phage Aehl complete genome."  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
RW [2]  
RP SEQUENCE FROM N.A.  
RA Petrov V., Nolan J., Karam J.D.;  
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY266303; AAQ17796.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 1211 AA; 137634 MW; 6447E705BB2A9A88 CRC64;

Query Match 43.98; Score 47; DB 2; Length 1211;  
Best Local Similarity 39.18; Pred. No. 1.9e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 4;

Query 1 RSDYKFYEDANG----TRDHKKG 19  
DB 773 KNGFKFFNAENGKSISVRDHKKG 795

RESULT 23

Q6CVA3  
ID Q6CVA3 PRELIMINARY; PRT; 1467 AA.  
AC Q6CVA3;  
DT 01-OCT-2004 (TRENBLrel. 28, Created)  
DT 01-OCT-2004 (TRENBLrel. 28, Last sequence update)  
DT 01-OCT-2004 (TRENBLrel. 28, Last annotation update)  
DE Similar to sp|P3202 Saccharomyces cerevisiae YKL010c SOS1.  
GN ORFNAME=KLLA0B13585g;  
OS Kluyveromyces lactis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
ON NCBI\_TaxID=28985;  
OX [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=NRRL Y-1140;  
RG GENOLEVURES;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talla E.,



RA Goffard N., Prangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
RA Boissame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.M., Nikolski M., Ostas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.F., Straub M.D., Suleau A.,  
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Winkler P., Souciet J.L.,  
RT "Genome evolution in yeasts."  
RL Nature 430:35-44(2004).  
RN [2]  
RC SEQUENCE FROM N.A.  
RP STRAIN=NRRL Y-1140;  
RA Genoscope;  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; CR382122; CAH02529.1; --  
SQ SEQUENCE 1467 AA; 165727 MW; F47CEDDD4F914998 CRC64;  
Query Match 43.9%; Score 47; DB 2; Length 1467;  
Best Local Similarity 61.5%; Pred. No. 2.4e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Oy 3 DYKPYEDANGTRD 15  
Db 9 DYIIYEDANGSHE 21  
RESULT 24  
SPAP STRMU  
ID -SPAP STRMU STANDARD; PRT; 1562 AA.  
AC P23504;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Cell surface antigen I/II precursor.  
GN Name=spap; OrderedLocusNames=SMU.610;  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=NG5 / Serotype c;  
RX MEDLINE=90076473; PubMed=2687020;  
RA Kelly C., Evans P., Bergmeier L., Lee S.F., Proguleske-Fox A.,  
RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;  
RT "Sequence analysis of the cloned streptococcal cell surface antigen  
I/II."  
RL FEBS Lett. 258:127-132(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NG5 / Serotype c;  
RX MEDLINE=91207143; PubMed=1982405;  
RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,  
RA Lee S.F., Bleiweis A.S., Lehner T.;  
RT "Sequencing and characterization of the 185 kDa cell surface antigen  
of Streptococcus mutans."  
RL Arch. Oral Biol. 35:338-385(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
RA Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
pathogen."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -1- FUNCTION: Surface protein antigen implicated in dental caries.  
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (Potential).  
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>  
CC or send an email to [license@sib-eib.ch](mailto:license@sib-eib.ch)).  
CC -----  
DR EMBL; X17390; CAA35253.1; --  
DR EMBL; AE014905; AAN58348.1; --  
DR PIR; S06839; S06839.  
DR HSP; P11657; LQMM.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR009578; Strep\_SA\_rep.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF06696; Strep\_SA\_rep; 7.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Antigen; Cell wall; Complete proteome; Dental caries;  
KW Direct protein sequencing; Peptidoglycan-anchor; Repeat; Signal.  
FT SIGNAL 1 38  
FT CHAIN 39 ? Cell surface antigen I.  
FT CHAIN 998 1532 Cell surface antigen II (Probable).  
FT PROPEP 1533 1562 Removed by sortase (Potential).  
FT DOMAIN 60 551 Helical (Potential).  
FT DOMAIN 220 465 3 X tandem repeats, Ala-rich.  
FT DOMAIN 848 964 3 X tandem repeats, Pro-rich.  
FT SITE 1529 1533 Pentaglycyl murein peptidoglycan amidated  
FT MOTIF 1532 1532 threonine (Potential).  
FT CONFLICT 73 73 E -> Q (in Ref. 1 and 2).  
FT CONFLICT 79 90 NQAGETNGSIPV -> TKLSRQMVTI (in Ref. 1  
and 2).  
FT CONFLICT 179 179 E -> G (in Ref. 1 and 2).  
FT CONFLICT 183 183 A -> V (in Ref. 1 and 2).  
FT CONFLICT 227 227 A -> S (in Ref. 1 and 2).  
FT CONFLICT 824 824 V -> L (in Ref. 1 and 2).  
FT CONFLICT 878 878 N -> K (in Ref. 1 and 2).  
FT CONFLICT 984 984 V -> I (in Ref. 1 and 2).  
FT CONFLICT 1010 1010 I -> V (in Ref. 1 and 2).  
FT CONFLICT 1069 1069 T -> A (in Ref. 1 and 2).  
FT CONFLICT 1120 1120 T -> S (in Ref. 1 and 2).  
FT CONFLICT 1201 1201 K -> Q (in Ref. 1 and 2).  
FT CONFLICT 1241 1241 N -> S (in Ref. 1 and 2).  
FT CONFLICT 1307 1307 I -> V (in Ref. 1 and 2).  
FT CONFLICT 1323 1323 I -> V (in Ref. 1 and 2).  
FT CONFLICT 1371 1371 D -> N (in Ref. 1 and 2).  
FT CONFLICT 1406 1410 FKDGS -> LKNGV (in Ref. 1 and 2).  
FT CONFLICT 1416 1416 A -> T (in Ref. 1 and 2).  
FT CONFLICT 1429 1429 A -> T (in Ref. 1 and 2).  
FT CONFLICT 1494 1494 T -> A (in Ref. 1 and 2).  
FT CONFLICT 1512 1512 N -> I (in Ref. 1 and 2).  
FT CONFLICT 1527 1527 E -> K (in Ref. 1 and 2).  
SQ SEQUENCE 1562 AA; 169971 MW; 298B244E7A95F5D7 CRC64;  
Query Match 43.9%; Score 47; DB 1; Length 1562;  
Best Local Similarity 47.1%; Pred. No. 2.5e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
Oy 3 DYKPYEDANGTRDHKKG 19  
Db 1378 EYNFYDDYDQTDGHTG 1394  
RESULT 25  
PAC STRMU  
ID -PAC STRMU STANDARD; PRT; 1565 AA.  
AC P11657;



```

SQ SEQUENCE 1565 AA; 170781 MW; 4C3B05C809D0C32A CRC64;
Query Match 43.9%; Score 47; DB 2; Length 1565;
Best Local Similarity 47.1%; Pred. No. 2.5e+02;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKPYEDANGTRDHKKG 19
Db 1381 EYNFYDDYDQGDHYTG 1397

RESULT 27
Q6HA27 PRELIMINARY; PRT; 698 AA.
AC Q6HA27
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Polyl oligopeptidase (EC 3.4.21.26).
GN Name=pop;
OS Trypanosoma brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RP SEQUENCE FROM N.A.
RA Bastos I.M.D., Santana J.M., Grellier P.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ496456; CAD42967.1; -
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR002470; Peptidase S9A.
DR InterPro; IPR004106; Peptidase S9A_N.
DR InterPro; IPR002471; Pept_S9_AS.
DR InterPro; IPR000379; Ser_estr.
DR Pfam; PF0326; Peptidase S9; 1.
DR PRINTS; PR02897; Peptidase S9 N; 1.
DR PRINTS; PR00862; PROLIGOPTASE.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 698 AA; 77597 MW; A969775872E45910 CRC64;

Query Match 43.5%; Score 46.5; DB 2; Length 698;
Best Local Similarity 43.5%; Pred. No. 1.3e+02;
Matches 10; Conservative 4; Mismatches 4; Indels 5; Gaps 1;

Qy 2 SDYKFYEDANGTR-----DHKKG 19
Db 433 TEQKFYNSADGTRIPMFIIHRKG 455

RESULT 28
Q6NWJ2 PRELIMINARY; PRT; 180 AA.
AC Q6NWJ2
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein zgc:85675.
GN Name=zgc:85675;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein zgc:85675.
GN Name=zgc:85675;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```

```

RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC067570; AAH67570.1; -
DR InterPro; IPR007653; SPC22.
DR InterPro; IPR010916; TONB_Box_N.
DR Pfam; PF04573; SPC22; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 180 AA; 20253 MW; 50FPD3C2B058A5AD CRC64;

Query Match 43.0%; Score 46; DB 2; Length 180;
Best Local Similarity 47.1%; Pred. No. 36;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHK 17
Db 124 KSKYFFFDGNGLRANK 140

RESULT 29
AAH67570 PRELIMINARY; PRT; 180 AA.
ID AAH67570
AC AAH67570;
DT 24-MAY-2004 (TrEMBLrel. 27, Created)
DT 24-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 24-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein zgc:85675.
GN ZGC:85675.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";

```





```

DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR007110; Ig-like.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
KW Complete proteome.
SQ SEQUENCE 1091 AA; 121698 MW; 55FC3969F472DCDE CRC64;

Query Match 43.0%; Score 46; DB 2; Length 1091;
Best Local Similarity 63.6%; Pred. No. 2.5e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANG 12
|:|:|:|:|
Db 864 SEYSFYDDVNG 874

RESULT 36
Q71WY0 PRELIMINARY; PRT; 1091 AA.
AC Q71WY0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Glycosyl hydrolase, family 31.
GN OrderedLocusNames=LMOF2365_2419;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=265669;
RN [1]
RP PubMed=15115801; DOI=10.1093/nar/gkh562;
RX Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Niernan W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Unlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytogenes reveal new insights into the
RT core genome components of this species.";
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AE017330; AAT05185.1; -.
KW Hydrolase.
SQ SEQUENCE 1091 AA; 121658 MW; 359049E0243DDB8B CRC64;

Query Match 43.0%; Score 46; DB 2; Length 1091;
Best Local Similarity 63.6%; Pred. No. 2.5e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANG 12
|:|:|:|:|
Db 864 SEYSFYDDVNG 874

RESULT 37
Q71WY0 PRELIMINARY; PRT; 1091 AA.
AC Q71WY0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Glycosyl hydrolase, family 31.
GN OrderedLocusNames=LMOF2365_2419;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=265669;
RN [1]
RP PubMed=15115801; DOI=10.1093/nar/gkh562;
RX Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Niernan W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Unlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytogenes reveal new insights into the
RT core genome components of this species.";
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AE017330; AAT05185.1; -.
GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR007110; Ig-like.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 1091 AA; 121658 MW; 359049E0243DDB8B CRC64;

Query Match 43.0%; Score 46; DB 2; Length 1091;
Best Local Similarity 63.6%; Pred. No. 2.5e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANG 12
|:|:|:|:|
Db 864 SEYSFYDDVNG 874

RESULT 38
Q71WY0 PRELIMINARY; PRT; 1091 AA.
AC Q71WY0;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hypothetical 39.6 kDa protein in GTR2-KRE11 intergenic region.
GN OrderedLocusNames=YGR165W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72950; CAA97189.1; -.
DR PIR; S64476; S64476.
DR GERMOnline; 141477; -.
DR SGD; S0003397; YGR165W.
DR GO; GO:0005763; C:mitochondrial small ribosomal subunit; IPI.
DR GO; GO:0003735; F:structural constituent of ribosome; IPI.
DR GO; GO:0009060; P:aerobic respiration; IEP.
DR GO; GO:0006412; P:protein biosynthesis; IC.
KW Hypothetical protein.
SQ SEQUENCE 345 AA; 39575 MW; 4E363E30F5056329 CRC64;

Query Match 42.5%; Score 45.5; DB 1; Length 345;

```

Best Local Similarity 40.7%; Pred. No. 87;  
Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Qy 1 RSDYKF-----YDANGTRDHKK 18

Db 304 RSQYKFTWAKVGKGYRSGNRDNKK 330

RESULT 39

Q8KAK9

ID Q8KAK9 PRELIMINARY; PRT; 381 AA.

AC Q8KAK9;

DT 01-OCT-2002 (TRENBLrel. 22, Created)

DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Hypothetical protein CT2147.

GN OrderedLocusNames=CT2147;

OS Chlorobium tepidum.

OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;

OC Chlorobaculum.

OX NCBI\_TaxID=1097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TLS / ATCC 49652 / DSM 12025;

RX MEDLINE=22103685; PubMed=12093901; DOI=10.1073/pnas.132181499;

RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,

RA Dodson R.J., DeBoy R.T., Gwinn M.L., Nelson W.C., Haft D.H.,

RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.F., Yang F.,

RA Holt I.E., Umayam L.A., Mason T.M., Brenner M., Shea T.P.,

RA Parksey D.S., Niernan W.C., Feldblyum T.V., Hansen C.L., Craven M.B.,

RA Radune D., Vamathevan J.J., Khouri H.M., White O., Gruber T.M.,

RA Ketchum K.A., Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;

RT "The complete genome sequence of Chlorobium tepidum TLS, a

RL photosynthetic, anaerobic, green-sulfur bacterium.";

DR EMBL; AE012963; AAM73363.1; --

DR TIGR; CT2147; --

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 381 AA; 42377 MW; 2CDDDEF72C48C479 CRC64;

Query Match 42.5%; Score 45.5; DB 2; Length 381;

Best Local Similarity 62.5%; Pred. No. 97;

Matches 10; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

Qy 1 RSDYKFYEDANGTRDH 16

Db 135 RSEVKLYE--GTRDH 147

RESULT 40

Q8B488

ID Q8B488 PRELIMINARY; PRT; 890 AA.

AC Q8B488;

DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE NS5 protein (Fragment).

GN Name=NS5;

OS Iquape virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus.

OX NCBI\_TaxID=64308;

RN [1]

RP SEQUENCE FROM N.A.

RA Baleotti F.G., Moreli M.L., Figueiredo L.T.M.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY167441; AAN77511.1; --

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR InterPro; IPR00208; Flavi NS5.

DR InterPro; IPR002877; RtmJFtsJ\_mtfase.

DR Pfam; PF00972; Flavi\_NS5; 1.

DR Pfam; PF01728; FtsJ; 1.

FT NON TER 1 1  
FT NON\_TER 890 890  
SQ SEQUENCE 890 AA; 100683 MW; 199C3122BDF4382B CRC64;

Query Match 42.5%; Score 45.5; DB 2; Length 890;

Best Local Similarity 64.7%; Pred. No. 2.4e+02;

Matches 11; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

Qy 5 KFYEDAN---GTRDHKK 18

Db 248 KFEEDANLSSGTRAHSK 264

Search completed: November 24, 2004, 09:29:01

Job time : 90.3587 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:44 ; Search time 15.6957 Seconds  
(without alignments)  
116.473 Million cell updates/sec

Title: US-09-719-379A-5  
Perfect score: 107  
Sequence: 1 RSDYKFYEDANGTRDHHKG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 263416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 47.5  | 44.4        | 366    | 2 E85729 | probable outer mem  |
| 2          | 47.5  | 44.4        | 366    | 2 D90888 | putative outer memb |
| 3          | 47.5  | 44.4        | 997    | 2 T28872 | hypothetical prote  |
| 4          | 47    | 43.9        | 304    | 2 G86143 | probable zinc fing  |
| 5          | 47    | 43.9        | 886    | 2 T10890 | cysteine proteinas  |
| 6          | 47    | 43.9        | 1561   | 1 S06839 | surface antigen sp  |
| 7          | 47    | 43.9        | 1565   | 2 S04729 | surface antigen pa  |
| 8          | 46    | 43.0        | 1090   | 2 AG1749 | glycosidase homolo  |
| 9          | 46    | 43.0        | 1091   | 2 AF1380 | glycosidase homolo  |
| 10         | 45.5  | 42.5        | 345    | 2 S64476 | hypothetical prote  |
| 11         | 45    | 42.1        | 339    | 2 C90126 | DNA-directed RNA p  |
| 12         | 44    | 41.1        | 142    | 2 G84201 | diadenosine tetrap  |
| 13         | 44    | 41.1        | 303    | 2 G96638 | protein t1F9.22 [i  |
| 14         | 43    | 40.2        | 326    | 2 T20776 | hypothetical prote  |
| 15         | 43    | 40.2        | 329    | 2 G88864 | protein T04A11.11   |
| 16         | 43    | 40.2        | 329    | 2 S27887 | Gene Tc3 protein -  |
| 17         | 43    | 40.2        | 329    | 2 E87901 | protein Jc247.4 [i  |
| 18         | 43    | 40.2        | 329    | 2 E87969 | protein Y47H9C.3 [  |
| 19         | 43    | 40.2        | 329    | 2 G88210 | protein R10H1.3 [i  |
| 20         | 43    | 40.2        | 329    | 2 H89382 | protein 2K218.2 [i  |
| 21         | 43    | 40.2        | 329    | 2 F87991 | protein V6B3B.8 [i  |
| 22         | 43    | 40.2        | 329    | 2 H87720 | protein F56A5.3 [i  |
| 23         | 43    | 40.2        | 329    | 2 D86198 | protein T02G5.5 [i  |
| 24         | 43    | 40.2        | 329    | 2 A88892 | protein Y45F10D.1   |
| 25         | 43    | 40.2        | 786    | 2 C86406 | 88.6K hypothetical  |
| 26         | 43    | 40.2        | 1394   | 2 A29637 | position-specific   |
| 27         | 43    | 40.2        | 2369   | 2 T18472 | hypothetical prote  |
| 28         | 42.5  | 39.7        | 995    | 2 T22942 | hypothetical prote  |
| 29         | 42    | 39.3        | 298    | 2 D86246 | hypothetical prote  |

|    |      |      |      |          |                    |
|----|------|------|------|----------|--------------------|
| 30 | 42   | 39.3 | 301  | 2 T29384 | hypothetical prote |
| 31 | 42   | 39.3 | 305  | 2 G96901 | D-3-phosphoglycera |
| 32 | 42   | 39.3 | 361  | 2 F86505 | tRNA (guanine N-1) |
| 33 | 42   | 39.3 | 361  | 2 C72117 | tRNA (guanine-N1)- |
| 34 | 42   | 39.3 | 423  | 2 A83691 | hypothetical prote |
| 35 | 42   | 39.3 | 801  | 2 H83737 | glucosidase BH0704 |
| 36 | 41.5 | 38.8 | 1130 | 2 T30251 | repetin - mouse    |
| 37 | 41   | 38.3 | 180  | 2 S22412 | signal peptidase ( |
| 38 | 41   | 38.3 | 202  | 2 H69465 | hypothetical prote |
| 39 | 41   | 38.3 | 264  | 2 G82939 | conserved hypothet |
| 40 | 41   | 38.3 | 269  | 2 S51815 | hypothetical prote |
| 41 | 41   | 38.3 | 298  | 2 T28233 | ORF MSV072 hypothe |
| 42 | 41   | 38.3 | 307  | 2 D89943 | hypothetical prote |
| 43 | 41   | 38.3 | 361  | 2 T30402 | hypothetical prote |
| 44 | 41   | 38.3 | 519  | 2 E90086 | t-complex protein  |
| 45 | 41   | 38.3 | 554  | 2 A70120 | hypothetical prote |

#### ALIGNMENTS

##### RESULT 1

E85729

probable outer membrane porin protein Z2239 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004

C:Accession: E85729

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouasis, K.; Apodaca,

Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85729

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-366 <STO>

A:Cross-references: UNIPROT:Q8XAS0; GB:AE005174; NID:gl2515212; PIDN:AAG56297.1; GSPDB:G

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z2239

C:Superfamily: outer membrane protein phoE

Query Match 44.4%; Score 47.5; DB 2; Length 366;

Best Local Similarity 62.5%; Pred. No. 13;

Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 3 DYKFYED-ANGTRDHHK 17

| : | : | : | : | : |

Db 80 DFKGYNDEANGSRDNK 95

##### RESULT 2

D90888

putative outer membrane porin protein Ecs2076 [imported] - Escherichia coli (strain O157:

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C:Accession: D90888

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 9, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90888

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-366 <HAY>

A:Cross-references: UNIPROT:Q8XAS0; GB:BA000007; PIDN:BA835499.1; PID:gl3361542; GSPDB:G

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: Ecs2076

C:Superfamily: outer membrane protein phoE

Query Match 44.4%; Score 47.5; DB 2; Length 366;

Best Local Similarity 62.5%; Pred. No. 13;  
Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 3 DYKFVED-ANGTRDHK 17  
| : | | | | | : | : | : | : | : |  
Db 80 DFKGYDEANGSRDNK 95

RESULT 3  
T28872  
hypothetical protein R04E5.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28872  
R:Miller, N.  
submitted to the EMBL Data Library, December 1995  
A:Description: The sequence of C. elegans comid R04E5.  
A:Reference number: Z20535  
A:Accession: T28872  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-997 <MIL>  
A:Cross-references: UNIPROT:Q9GYL4; EMBL:U41538; PIDN:RAC48181.1; GSPDB:GNO00028; CESP:R04E5  
A:Experimental source: strain Bristol N2; clone R04E5  
C:Genetics:  
A:Gene: CESP:R04E5.8  
A:Map position: X  
A:Introns: 216/3; 234/3; 255/2; 305/1; 669/3; 891/1; 986/2

Query Match 44.4%; Score 47.5; DB 2; Length 997;  
Best Local Similarity 62.5%; Pred. No. 33;  
Matches 10; Conservative 2; Mismatches 1; Indels 3; Gaps 1;

Qy 2 SDYKFY---EDANGTR 14  
| : | : | | | | : | : | : | : | : |  
Db 289 SGYRFYMGEDANGTK 304

RESULT 4  
G86143  
Probable zinc finger protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G86143  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.B.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G86143  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-304 <STO>  
A:Cross-references: UNIPROT:Q9LNI9; GB:AE005172; NID:9965151; PIDN:AAF97335.1; GSPDB:GN  
C:Genetics:  
A:Map position: 1

Query Match 43.9%; Score 47; DB 2; Length 304;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANGTRDHKKG 19  
| : | : | : | : | : | : | : | : | : | : |  
Db 151 SDEKLYGIHYTDHKG 168

RESULT 5

T10990  
cysteine proteinase (EC 3.4.22.-) / hemagglutinin protein - Porphyromonas gingivalis  
C:Species: Porphyromonas gingivalis  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T10890  
R:Madden, T.E.; Clark, V.L.; Kuramitsu, H.K.  
Infect. Immun. 63, 238-247, 1995  
A:Title: Revised sequence of the Porphyromonas gingivalis prtT cysteine protease/hemaggl  
A:Reference number: Z17199; MUID:95105001; PMID:7806362  
A:Accession: T10890  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-886 <MAD>  
A:Cross-references: UNIPROT:Q53481; EMBL:S75942; NID:G913136; PID:G913137  
C:Genetics:  
A:Gene: prtT  
C:Keywords: cysteine proteinase; hemagglutinin; hydrolase

Query Match 43.9%; Score 47; DB 2; Length 886;  
Best Local Similarity 41.2%; Pred. No. 35;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DYKFYEDANGTRDHKKG 19  
| : | : | : | : | : | : | : | : | : | : |  
Db 223 EYDYDDMTGTHYSG 239

RESULT 6  
S06839  
surface antigen spap precursor - Streptococcus mutans  
N:Alternate names: antigen I/II  
C:Species: Streptococcus mutans  
C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 09-Jul-2004  
C:Accession: S06839; A60339; A60661  
R:Kelly, C.; Evans, P.; Bergmeier, L.; Lee, S.F.; Proguiske-Fox, A.; Harris, A.C.; Aitke  
FEBS Lett. 258, 127-132, 1989  
A:Title: Sequence analysis of the cloned streptococcal cell surface antigen I/II.  
A:Reference number: S06839; MUID:90076473; PMID:2687020  
A:Accession: S06839  
A:Molecule type: DNA  
A:Residues: 1-1561 <KEL>  
A:Cross-references: UNIPROT:P23504; EMBL:X17390; NID:947266; PIDN:CAA35253.1; PID:947267  
A:Experimental source: serotype c  
A:Note: parts of this sequence, including the amino ends of the mature proteins, were co  
R:Kelly, C.; Evans, P.; Ma, J.K.C.; Bergmeier, L.A.; Taylor, W.; Brady, L.J.; Lee, S.F.;  
Arch. Oral Biol. 35(Suppl.), 33S-38S, 1990  
A:Title: Sequencing and characterization of the 185 kDa cell surface antigen of Streptoc  
A:Reference number: A60661; MUID:91207143; PMID:1982405  
A:Contents: annotation  
R:Ma, J.K.C.; Kelly, C.G.; Munro, G.; Whiley, R.A.; Lehner, T.  
Infect. Immun. 59, 2686-2694, 1991  
A:Title: Conservation of the gene encoding streptococcal antigen I/II in oral streptococ  
A:Reference number: A60339; MUID:91310321; PMID:1855988  
A:Accession: A60339  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1084-1189 <MAA>  
C:Genetics:  
A:Gene: spap  
C:Function:  
A:Description: probably plays a role in adherence to the tooth surface.  
C:Superfamily: surface antigen spap  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-38/Domain: signal sequence #status predicted <SIG>  
F:39-1561/Product: surface antigen spap.I #status predicted <MATI>  
F:169-193/Domain: spap alanine-rich repeat <AR1>  
F:194-218/Domain: spap alanine-rich repeat <AR2>  
F:219-243/Domain: spap alanine-rich repeat <AR3>  
F:244-268/Domain: spap alanine-rich repeat <AR4>  
F:276-300/Domain: spap alanine-rich repeat <AR5>  
F:301-325/Domain: spap alanine-rich repeat <AR6>  
F:326-350/Domain: spap alanine-rich repeat <AR7>  
F:358-382/Domain: spap alanine-rich repeat <AR8>

F:383-407/Domain: spap alanine-rich repeat <AR9>  
 F:408-426/Domain: spap alanine-rich repeat <AR10>  
 F:440-464/Domain: spap alanine-rich repeat <AR11>  
 F:847-885/Domain: spap proline-rich repeat <PR1>  
 F:886-924/Domain: spap proline-rich repeat <PR2>  
 F:925-963/Domain: spap proline-rich repeat <PR3>  
 F:997-1561/Product: surface antigen spap.II #status predicted <MAT2>

Query Match 43.9%; Score 47; DB 1; Length 1561;  
 Best Local Similarity 47.1%; Pred. No. 60;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKPYEDANGTRDHKKG 19  
 Db 1377 EYFYDDYDQGDHYTG 1393

## RESULT 7

S04729  
 surface antigen pac precursor - Streptococcus mutans  
 C:Species: Streptococcus mutans  
 C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 09-Jul-2004  
 C:Accession: S04729  
 R:Okahashi, N.; Sasakawa, C.; Yoshikawa, M.; Hamada, S.; Koga, T.  
 Mol. Microbiol. 3, 673-678, 1989  
 A>Title: Molecular characterization of a surface protein antigen gene from serotype c S04729  
 A:Reference number: S04729; MUID:89343654; PMID:2761390  
 A:Accession: S04729  
 A:Molecule type: DNA  
 A:Residues: 1-1565 <OKA>  
 A:Cross-references: UNIPROT:P11657; EMBL:X14490; NID:G47247; PIDN:CAA32652.1; PID:G47248  
 A>Note: part of this sequence, including the amino end of the mature protein, was confirmed by sequencing  
 C:Genetics:  
 A:Gene: pac

Query Match 43.9%; Score 47; DB 2; Length 1565;  
 Best Local Similarity 47.1%; Pred. No. 60;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 Qy 3 DYKPYEDANGTRDHKKG 19  
 Db 1381 EYFYDDYDQGDHYTG 1397

## RESULT 8

AG1749  
 glycosidase homolog lin2540 [imported] - Listeria innocua (strain Clip11262)  
 C:Species: Listeria innocua  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AG1749  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AG1749  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1090 <GLA>  
 A:Cross-references: UNIPROT:Q928J2; GB:AL592022; PIDN:CAC97767.1; PID:gl6415062; GSPDB:C90126  
 A:Experimental source: strain Clip11262  
 C:Genetics:  
 A:Gene: lin2540

Query Match 43.0%; Score 46; DB 2; Length 1090;  
 Best Local Similarity 63.6%; Pred. No. 60;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 SDYKPYEDANG 12  
 Db 864 SEYSFYDDVNG 874

## RESULT 9

AF1380  
 glycosidase homolog lmo2446 [imported] - Listeria monocytogenes (strain EGD-e)  
 C:Species: Listeria monocytogenes  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AF1380  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AF1380  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1091 <GLA>  
 A:Cross-references: UNIPROT:Q8Y4J2; GB:NC\_003210; PIDN:CAD00524.1; PID:gl6411934; GSPDB:C90126  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo2446

Query Match 43.0%; Score 46; DB 2; Length 1091;  
 Best Local Similarity 63.6%; Pred. No. 61;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SDYKPYEDANG 12  
 Db 864 SEYSFYDDVNG 874

## RESULT 10

S64476  
 hypothetical protein YGR165w - Yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein G7050  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 09-Jul-2004  
 C:Accession: S64476  
 R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.  
 submitted to the Protein Sequence Database, May 1996  
 A:Reference number: S64071  
 A:Accession: S64476  
 A:Molecule type: DNA  
 A:Residues: 1-345 <RIE>  
 A:Cross-references: UNIPROT:P53292; EMBL:Z72950; NID:gl323289; PID:G243551; PID:gl323290  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: MIP5:YGR165w  
 A:Cross-references: SGD:S0003397  
 A:Map position: 7R

Query Match 42.5%; Score 45.5; DB 2; Length 345;  
 Best Local Similarity 40.7%; Pred. No. 24;  
 Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Qy 1 RSDYKF-----YEDANGTRDHKK 18  
 Db 304 RSQYKFTNAKVGKVGRYGSGNRDNKK 330

## RESULT 11

C90126  
 DNA-directed RNA polymerase 40k chain [imported] - Guillardia theta nucleomorph  
 C:Species: nucleomorph Guillardia theta  
 A>Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: C90126  
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re;  
Nature 410, 1091-1096, 2001  
A;Title: The highly reduced genome of an enslaved algal nucleus.  
A;Reference number: A99082; MUID:11323671; PMID:11323671  
A;Accession: C90126  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-339 <DOU>  
A;Cross-references: UNIPROT:Q98579; GB:AF083031; NID:gi13794326; PIDN:AAK39703.1; GSPDB:C  
C;Genetics:  
A;Gene: tpa5  
A;Map position: 3  
A;Genome: nucleomorph  
C;Superfamily: Saccharomyces cerevisiae DNA-directed RNA polymerase 40K chain  
C;Keywords: nucleomorph

Query Match 42.1%; Score 45; DB 2; Length 339;  
Best Local Similarity 43.8%; Pred. NO. 28;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKFVEDANGTRDHKKG 19  
||::|:|:|  
Db 2 YKIFDTINLNKDNKKG 17  
||::|:|:|

RESULT 12  
G84201  
diadenosine tetraphosphate pyrophosphohydrolase [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: G84201  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Barquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Lethauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: G84201  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-142 <STO>  
A;Cross-references: UNIPROT:Q9HS29; GB:AE004437; NID:gi0580041; PIDN:AAG18979.1; GSPDB:C  
C;Genetics:  
A;Gene: apa

Query Match 41.1%; Score 44; DB 2; Length 142;  
Best Local Similarity 47.1%; Pred. NO. 17;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKFVEDANGTRDHK 17  
|:|:|:|:|:|  
Db 69 RDEYDYVFEANGDRHK 85  
|:|:|:|:|:|

RESULT 13  
G96638  
protein TlP9.22 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: G96638  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G96638  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-303 <STO>  
A;Cross-references: UNIPROT:O64791; GB:AE005173; NID:g3056601; PIDN:AAC13912.1; GSPDB:GN  
C;Genetics:  
A;Gene: TlP9.22  
A;Map position: 1  
C;Superfamily: syntaxin

Query Match 41.1%; Score 44; DB 2; Length 303;  
Best Local Similarity 46.7%; Pred. NO. 36;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 5 KFVEDANGTRDHKKG 19  
||:|:|:|:|:|  
Db 35 KPFEDVENVDNMKG 49  
||:|:|:|:|:|

RESULT 14  
T20776  
hypothetical protein F11D11.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T20776  
R;Mortimore, B.; Basham, V.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19323  
A;Accession: T20776  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-326 <WIL>  
A;Cross-references: UNIPROT:O62151; EMBL:Z81500; PIDN:CAB04094.1; GSPDB:GN000023; CESP:F  
A;Experimental source: clone F11D11  
C;Genetics:  
A;Gene: CESP:F11D11.2  
A;Map position: 5  
A;Introns: 136/3  
C;Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 326;  
Best Local Similarity 37.5%; Pred. NO. 54;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFVED-----ANGTRDHKK 18  
|:|:|:|:|:|:|:|:|  
Db 220 RKDFRFQDDNATIHVSNSTRDYFK 243  
|:|:|:|:|:|:|:|:|

RESULT 15  
G88864  
protein T04A11.11 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C;Accession: G88864  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C\_ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: G88864  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-329 <STO>  
A;Cross-references: GB:chr\_IV; PIDN:CAB05614.1; PID:g3879394; GSPDB:GN000022; CESP:T04A11  
C;Genetics:  
A;Gene: T04A11.11  
A;Map position: 4  
C;Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 329;

Best Local Similarity 37.58; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

QY 1 RSDYKFYED-----ANGTRDHK 18  
| : | : | : | : | : | : |  
DB 223 RKDFRFOODNATIHVSNSTRDYFK 246

RESULT 16  
S27787  
gene Tc3 protein - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text change 09-Jul-2004  
C;Accession: S27787 #T19124; T16785; T19136; T27325; T26932  
R;Du, Z.; Ainscough, R.; Berks, M.; Craxton, A.; Coulson, A.; Dear, S.; Durbin, R.K.; Grainger, R.; Hegerl, R.; Krawinkel, U.; Lott, J.; Mowbray, J.; O'Neil, J.; Pridmore, D.; Rothermel, B.; Shownkeen, R.; Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, J.D.; White, S.  
submitted to the EMBL Data Library, October 1991  
A;Description: Sequence of the C. elegans cosmid B0303.  
A;Reference number: S27783  
A;Accession: S27787  
A;Molecule type: DNA  
A;Residues: 1-329 <DUZ>  
A;Cross-references: UNIPROT:P34257; EMBL:M77697; NID:g156188; PID:g156196  
R;Murray, J.; Rohlfing, T.; O'Neal, D.; Wilson, R.  
submitted to the EMBL Data Library, May 1998  
A;Description: The sequence of C. elegans cosmid F56A6.  
A;Reference number: Z21279  
A;Accession: T33080  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A;Molecule type: DNA  
A;Residues: 1-316,'F',318-329 <MUR>  
A;Cross-references: EMBL:AF067217; PIDN:AAC17016.1; GSPDB:GN00019; CESP:F56A6.3  
R;McMurray, A.  
submitted to the EMBL Data Library, September 1996  
A;Reference number: Z19076  
A;Accession: T19124  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A;Molecule type: DNA  
A;Residues: 1-316,'F',318-329 <WIL>  
A;Cross-references: EMBL:Z80213; PIDN:CAB02260.1; GSPDB:GN00022; CESP:F38A1.2  
A;Experimental source: clone C09E9  
R;Pauley, A.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid T02G5.  
A;Reference number: Z18577  
A;Accession: T16785  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A;Molecule type: DNA  
A;Residues: 1-316,'F',318-329 <PAU>  
A;Cross-references: EMBL:U41105; NID:g1086772; PID:g1086778; PIDN:AAA82401.1; CESP:T02G5  
R;Lloyd, C.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19489  
A;Accession: T21936  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A;Molecule type: DNA  
A;Residues: 1-316,'F',318-329 <W12>  
A;Cross-references: EMBL:Z81535; PIDN:CAB04359.1; GSPDB:GN00022; CESP:F38A1.2  
R;White, S.  
submitted to the EMBL Data Library, October 1998  
A;Reference number: Z20345  
A;Accession: T27325  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A;Molecule type: DNA  
A;Residues: 1-86,'F',88-316,'F',318-329 <W13>  
A;Cross-references: EMBL:AL032655; PIDN:CAA21724.1; GSPDB:GN00019; CESP:Y6B3B.8  
R;McMurray, A.  
submitted to the EMBL Data Library, January 1998  
A;Reference number: Z20288  
A;Accession: T26932

A;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-45,'L',47-316,'F',318-329 <Wt4>

A:Cross-references: EMBL:AL021492; PIDN:CAA16380.1; GSPDB:GN00022; CESP:Y45FI0D.1

A:Experimental source: clone Y45FI0D

C:Genetics: <GI>

A:Gene: CESP:P56A6.3; CESP:V6B3B.8; CESP:Y45FI0D.1

A:Map position: 1

A:Introns: 139/3

C:Genetics: <G2>

A:Gene: CESP:T02GS.5; CESP:F38A1.2

A:Map position: 4

A:Introns: 79/2; 139/3

C:Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

C:Keywords: DNA binding; nucleus

Query Match 40.2%; Score 43; DB 2; Length 329;  
Best Local Similarity 37.5%; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFYED-----ANGTRDHKK 18  
|::||:|:|::|  
Db 223 RKDFRFQQDNATIHVSNSRDYFK 246

RESULT 17

E87901

protein ZC247.4 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: E87901

R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998

A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_ele

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: E87901

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <STO>

A:Cross-references: UNIPROT:Q21913; GB:chr\_I; PIDN:CAB02311.1; PID:g3881428; GSPDB:GN000

C:Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 329;  
Best Local Similarity 37.5%; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFYED-----ANGTRDHKK 18  
|::||:|:|::|  
Db 223 RKDFRFQQDNATIHVSNSRDYFK 246

RESULT 18

E87969

protein Y47H9C.3 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: E87969

R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998

A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_ele

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: E87969

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <STO>

A:Cross-references: UNIPROT:Q21913; GB:chr\_I; PIDN:CAA21733.1; PID:g3881074; GSPDB:GN000



RESULT 23  
D88198  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D88198  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D88198  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <STO>  
A:Cross-references: UNIPROT:Q22104; GB:chr\_II; PIDN:AA82401.1; PID:gl086778; GSPDB:GN000  
C:Genetics:  
A:Gene: T02G5.5  
A:Map position: 2  
C:Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 329;  
Best Local Similarity 37.5%; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFYED-----ANGTRDHHK 18  
Db 223 RKDFRFOQDNATHVNSNSTRDYFK 246

RESULT 24  
A88892  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: A88892  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A88892  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <STO>  
A:Cross-references: UNIPROT:O62475; GB:chr\_IV; PIDN:CAAL6380.1; PID:g3880993; GSPDB:GN000  
C:Genetics:  
A:Gene: Y45P10D.1  
A:Map position: 4  
C:Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 329;  
Best Local Similarity 37.5%; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFYED-----ANGTRDHHK 18  
Db 223 RKDFRFOQDNATHVNSNSTRDYFK 246

RESULT 25  
C86406  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: C86406  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: C86406  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <STO>  
A:Cross-references: UNIPROT:O62475; GB:chr\_IV; PIDN:CAAL6380.1; PID:g3880993; GSPDB:GN000  
C:Genetics:  
A:Gene: Y45P10D.1  
A:Map position: 4  
C:Superfamily: Caenorhabditis transposon Tc1 hypothetical 32K protein

Query Match 40.2%; Score 43; DB 2; Length 329;  
Best Local Similarity 37.5%; Pred. No. 55;  
Matches 9; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

Qy 1 RSDYKFYED-----ANGTRDHHK 18  
Db 223 RKDFRFOQDNATHVNSNSTRDYFK 246

RESULT 26  
A29637  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: A29637  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A29637  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 201-301 <BRO>  
A:Gene: FlyBase:if  
A:Cross-references: FlyBase:FBgn0001250  
C:Keywords: cell adhesion; cytoskeleton; duplication; heterodimer; membrane protein

Query Match 40.2%; Score 43; DB 2; Length 1394;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANGTRDHHKK 19  
Db 774 SSYDFYWEANSTNLEKPG 791

RESULT 27  
T18472  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18472  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18472  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18472  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18472  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

Query Match 40.2%; Score 43; DB 2; Length 786;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKFYEDANGTRDHHK 18  
Db 98 RSKRKDRREENGARGDKK 115

RESULT 26  
A29637  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: A29637  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A29637  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 201-301 <BRO>  
A:Gene: FlyBase:if  
A:Cross-references: FlyBase:FBgn0001250  
C:Keywords: cell adhesion; cytoskeleton; duplication; heterodimer; membrane protein

Query Match 40.2%; Score 43; DB 2; Length 1394;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKFYEDANGTRDHHKK 19  
Db 774 SSYDFYWEANSTNLEKPG 791

RESULT 27  
T18472  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18472  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18472  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A;Residues: 1-2269 <LAW>  
A;Cross-references: UNIPROT:O77360; EMBL:AL008970; NID:e1407852; PID:e132566; PIDN:CAAL  
C;Genetics:  
A;Gene: C0440C  
A;Map position: 3

Query Match 40.2%; Score 43; DB 2; Length 2269;  
Best Local Similarity 50.0%; Pred. No. 3.5e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 YKFYEDANGTRDHHK 17  
|||||:|:|:|  
Db 182 YKFYDDKRNKRSNK 195

RESULT 28  
T22942  
hypothetical protein F58G11.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T22942  
R;Percy, C.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19640  
A;Accession: T22942  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-995 <WIL>  
A;Cross-references: UNIPROT:P09897; EMBL:Z81094; PIDN:CAB03153.1; GSPDB:GNO0023; CESP:PF  
A;Experimental source: clone F58G11  
C;Genetics:  
A;Gene: CESP:F58G11.2  
A;Map position: 5  
A;Introns: 79/1; 195/1; 642/1; 678/2; 796/1

Query Match 39.7%; Score 42.5; DB 2; Length 995;  
Best Local Similarity 42.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 RSDYKFYEDANGTRDHHK 19  
|||||:|:|:|  
Db 257 RNDYRSQQD---SRDHRSG 272

RESULT 29  
D86246  
hypothetical protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: D86246  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: D86246  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-298 <STO>  
A;Cross-references: UNIPROT:Q9SXB0; GB:AE005172; NID:95734739; PIDN:AAD50004.1; GSPDB:GN  
C;Genetics:  
A;Map position: 1  
C;Superfamily: syntaxin

Query Match 39.3%; Score 42; DB 2; Length 298;  
Best Local Similarity 46.7%; Pred. No. 71;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 KFYEDANGTRDHHK 19  
|||||:|:|:|  
Db 30 KFFEDVENVKDKMG 44

RESULT 30  
T29384  
hypothetical protein K08D10.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T29384  
R;Geisel, C.; Bradshaw, H.  
submitted to the EMBL Data Library, April 1996  
A;Description: The sequence of C. elegans cosmid K08D10.  
A;Reference number: Z20616  
A;Accession: T29384  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-301 <GEI>  
A;Cross-references: UNIPROT:Q21319; EMBL:U55857; PIDN:AAA98029.1; GSPDB:GN00022; CESP:K0  
A;Experimental source: strain Bristol N2; clone K08D10  
C;Genetics:  
A;Gene: CESP:K08D10.7  
A;Map position: 4  
A;Introns: 73/3; 193/3; 256/1  
C;Superfamily: Caenorhabditis elegans hypothetical protein K08D10.7

Query Match 39.3%; Score 42; DB 2; Length 301;  
Best Local Similarity 41.2%; Pred. No. 72;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 DYKFYEDANGTRDHHK 19  
|||||:|:|:|  
Db 276 DFTYFEDRNQDRHHGG 292

RESULT 31  
G96901  
D-3-phosphoglycerate dehydrogenase [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: G96901  
R;Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G96901  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-305 <KUR>  
A;Cross-references: UNIPROT:Q97N23; GB:AE001437; PIDN:AAK78002.1; PID:gl5022834; GSPDB:G  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0015  
C;Superfamily: phosphoglycerate dehydrogenase

Query Match 39.3%; Score 42; DB 2; Length 305;  
Best Local Similarity 35.7%; Pred. No. 72;  
Matches 10; Conservative 2; Mismatches 4; Indels 12; Gaps 1;

QY 4 YKFYEDANG-----TRDHHK 19  
|||||:|:|:|  
Db 184 YEFYDDINGLLREADFVSLHIPYDKKG 211

RESULT 32  
F86505  
tRNA (guanine N-1)-methyltransferase [imported] - Chlamydomonas reinhardtii (strain J138)  
C;Species: Chlamydomonas reinhardtii  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: F86505



R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Iida, T.; Nucleic Acids Res. 28, 2311-2316, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Accession: F86505  
 A:Reference: A86491; MUID:20330349; PMID:10871362  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-361 <STO>  
 A:Cross-references: UNIPROT:Q92964; GB:BA000008; NID:g9978491; PIDN:BA098328.1; GSPDB:GN000001  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: trmD

Query Match 39.3%; Score 42; DB 2; Length 361;  
 Best Local Similarity 50.0%; Pred. No. 85;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKFDVANGTRDHKK 18  
 |||||  
 Db 230 DHKFDDETTNRDHFK 245  
 |||||

RESULT 33  
 C72117  
 tRNA (guanine-N1)-methyltransferase CP0656 [imported] - Chlamydomonada pneumoniae (strain C-125)  
 C:Species: Chlamydomonada pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
 C:Accession: C72117; D81553  
 R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: C72117  
 A:Molecule type: DNA  
 A:Residues: 1-361 <ARN>  
 A:Cross-references: UNIPROT:Q92964; GB:AE001598; GB:AE001363; NID:g4376375; PIDN:AAD1827  
 A:Experimental source: strain C72117  
 R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.; Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255; PMID:10684935  
 A:Accession: D81553  
 A:Molecule type: DNA  
 A:Residues: 1-361 <REA>  
 A:Cross-references: GB:AR002223; GB:AE002161; NID:g7189565; PIDN:AAF38470.1; PID:g718957  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: trmD; CP0656

Query Match 39.3%; Score 42; DB 2; Length 361;  
 Best Local Similarity 50.0%; Pred. No. 85;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKFDVANGTRDHKK 18  
 |||||  
 Db 230 DHKFDDETTNRDHFK 245  
 |||||

RESULT 34  
 A83691  
 Hypothetical protein BH0329 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
 C:Accession: A83691  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y.; Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its relationship to other Bacillus species.  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: A83691  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-423 <STO>

A:Cross-references: UNIPROT:Q9KFF7; GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA0040  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0329.

Query Match 39.3%; Score 42; DB 2; Length 423;  
 Best Local Similarity 72.7%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKFFEDANG 12  
 |||||  
 Db 89 SDIGFYEDENG 99  
 |||||

RESULT 35  
 H83737  
 glucosidase BH0704 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
 C:Accession: H83737  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y.; Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its relationship to other Bacillus species.  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: H83737  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-801 <STO>  
 A:Cross-references: UNIPROT:Q9KE25; GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA0040  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0704  
 C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology

Query Match 39.3%; Score 42; DB 2; Length 801;  
 Best Local Similarity 43.8%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 4 YKFFEDANGTRDHKK 19  
 ::|||  
 Db 722 FRLYEDDGETNDYKDG 737  
 ::|||

RESULT 36  
 T30251  
 repetin - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T30251  
 R;Krieg, P.; Schuppler, M.; Koesters, R.; Mincheva, A.; Lichter, P.; Marks, F.; Genomics 43, 339-348, 1997  
 A:Title: Repetin (Rptn), a new member of the fused gene subgroup within the S100 gene family.  
 A:Reference number: Z20789; MUID:97422611; PMID:9268637  
 A:Accession: T30251  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1130 <KRI>  
 A:Cross-references: UNIPROT:P97347; EMBL:X99251; NID:g1806131; PIDN:CAAC7624.1; PID:g1806131  
 A:Experimental source: strain NMRI; clone 3031; skin papilloma  
 C:Genetics:  
 A:Map position: F2.1-2.2  
 A:Introns: 46/3  
 C:Keywords: calcium binding; EF hand; tandem repeat

Query Match 38.8%; Score 41.5; DB 2; Length 1130;  
 Best Local Similarity 41.7%; Pred. No. 3e+02;  
 Matches 10; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

Qy 3 DYKFFY-----EDANGTRDHK 17  
 |||||  
 Db 90 DSKFYGSRTSSQKEHDQEGTRSHK 113  
 |||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:14:14 ; Search time 16.7283 Seconds  
(without alignments)  
75.324 Million cell updates/sec

Title: US-09-719-379A-4  
Perfect score: 98  
Sequence: 1 RSDYKLYKNSTLKDLGE 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 49    | 50.0        | 885    | 2  | US-08-310-912A-2     |
| 2          | 49    | 50.0        | 885    | 3  | US-08-841-089-2      |
| 3          | 49    | 50.0        | 885    | 3  | US-09-301-085-2      |
| 4          | 49    | 50.0        | 885    | 5  | PCT-US95-04570-2     |
| 5          | 49    | 50.0        | 885    | 5  | PCT-US95-04589-2     |
| 6          | 49    | 50.0        | 907    | 3  | US-08-930-996A-7     |
| 7          | 49    | 50.0        | 907    | 2  | US-08-310-912A-142   |
| 8          | 49    | 50.0        | 909    | 3  | US-09-301-085-142    |
| 9          | 49    | 50.0        | 909    | 5  | PCT-US95-04589-142   |
| 10         | 45    | 45.9        | 582    | 4  | US-09-248-796A-16870 |
| 11         | 44.5  | 45.4        | 733    | 4  | US-09-328-352-5599   |
| 12         | 43    | 43.9        | 820    | 4  | US-09-248-796A-19901 |
| 13         | 42    | 42.9        | 511    | 3  | US-09-134-078-21     |
| 14         | 42    | 42.9        | 591    | 4  | US-09-248-796A-14458 |
| 15         | 41.5  | 42.3        | 176    | 4  | US-09-710-279-3262   |
| 16         | 41.5  | 42.3        | 357    | 3  | US-09-134-001C-3379  |
| 17         | 41.5  | 42.3        | 369    | 4  | US-09-809-665A-153   |
| 18         | 41.5  | 42.3        | 1464   | 2  | US-08-231-193A-11    |
| 19         | 41.5  | 42.3        | 1464   | 2  | US-08-486-273A-11    |
| 20         | 41.5  | 42.3        | 1464   | 3  | US-08-940-086A-11    |
| 21         | 41.5  | 42.3        | 1464   | 3  | US-08-436-332B-10    |
| 22         | 41.5  | 42.3        | 1464   | 3  | US-08-940-035A-11    |
| 23         | 41.5  | 42.3        | 1464   | 3  | US-08-935-105A-11    |
| 24         | 41.5  | 42.3        | 1464   | 4  | US-09-648-797-11     |
| 25         | 41.5  | 42.3        | 1464   | 4  | US-08-217-704C-2     |
| 26         | 41.5  | 42.3        | 1464   | 4  | US-09-386-123-11     |
| 27         | 41    | 41.8        | 126    | 4  | US-09-270-767-58231  |

28 41 41.8 182 4 US-09-634-238-370 Sequence 370, Appl  
29 41 41.8 245 4 US-09-198-452A-1091 Sequence 1091, Ap  
30 41 41.8 273 4 US-09-270-767-42905 Sequence 42905, A  
31 41 41.8 281 3 US-09-404-258-6 Sequence 6, Appli  
32 41 41.8 281 4 US-09-853-701-6 Sequence 6, Appli  
33 41 41.8 338 1 US-08-210-394-1 Sequence 1, Appli  
34 41 41.8 342 4 US-09-543-681A-4728 Sequence 4728, Ap  
35 41 41.8 351 4 US-09-198-452A-991 Sequence 991, App  
36 41 41.8 388 4 US-09-248-796A-18781 Sequence 18781, A  
37 41 41.8 476 3 US-09-306-593-9 Sequence 9, Appli  
38 41 41.8 509 3 US-09-134-078-18 Sequence 18, Appli  
39 41 41.8 509 4 US-10-039-293A-2 Sequence 2, Appli  
40 41 41.8 579 4 US-09-540-236-2373 Sequence 2373, Ap  
41 41 41.8 691 1 US-08-064-174-2 Sequence 2, Appli  
42 41 41.8 691 1 US-08-066-167-4 Sequence 4, Appli  
43 41 41.8 691 2 US-08-449-733-2 Sequence 2, Appli  
44 41 41.8 692 3 US-08-448-194-62 Sequence 62, Appli  
45 41 41.8 692 3 US-08-867-921-62 Sequence 62, Appli

#### ALIGNMENTS

RESULT 1  
US-08-310-912A-2  
; Sequence 2, Application US/08310912A  
; Patent No. 5981730  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Staskawicz, Brian J.  
; APPLICANT: Brent, Andrew F.  
; APPLICANT: Dahlbeck, Douglas  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kunkel, Barbara N.  
; APPLICANT: Mindrinos, Michael N.  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: RES2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 208  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2904  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/310,912A  
; FILING DATE: September 22, 1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/227,360  
; FILING DATE: April 13, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lech, Karen F.  
; REGISTRATION NUMBER: 35,238  
; REFERENCE/DOCKET NUMBER: 00786/254001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 100254  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 885 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

## US-08-310-912A-2

Query Match 50.0%; Score 49; DB 2; Length 885;  
Best Local Similarity 55.6%; Pred. No. 8.7;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGE 19  
:|||||:|:|:|:  
Db 103 ADYKLCCKVSAILKSIGE 120

## RESULT 2

US-08-841-089-2  
; Sequence 2, Application US/08841089  
; Patent No. 6127607  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Staskawicz, Brian J.  
; APPLICANT: Brent, Andrew F.  
; APPLICANT: Dahlbeck, Douglas  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kunkel, Barbara N.  
; APPLICANT: Mindrinos, Michael N.  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2904  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/841,089  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,360  
; FILING DATE: 13-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/230001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 100254  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 885 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-841-089-2

Query Match 50.0%; Score 49; DB 3; Length 885;  
Best Local Similarity 55.6%; Pred. No. 8.7;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGE 19  
:|||||:|:|:|:  
Db 103 ADYKLCCKVSAILKSIGE 120

## RESULT 3

US-09-301-085-2

; Sequence 2, Application US/09301085  
; Patent No. 6262248  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Staskawicz, Brian J.  
; APPLICANT: Brent, Andrew F.  
; APPLICANT: Dahlbeck, Douglas  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kunkel, Barbara N.  
; APPLICANT: Mindrinos, Michael N.  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND  
; TITLE OF INVENTION: DETECTION METHODS  
; FILE REFERENCE: 00786/254002  
; CURRENT APPLICATION NUMBER: US/09/301,085  
; CURRENT FILING DATE: 1999-04-28  
; EARLIER APPLICATION NUMBER: 08/310,912  
; EARLIER FILING DATE: 1994-09-22  
; EARLIER APPLICATION NUMBER: 08/227,360  
; EARLIER FILING DATE: 1994-04-13  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 885  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; US-09-301-085-2

Query Match 50.0%; Score 49; DB 3; Length 885;  
Best Local Similarity 55.6%; Pred. No. 8.7;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGE 19  
:|||||:|:|:|:  
Db 103 ADYKLCCKVSAILKSIGE 120

## RESULT 4

PCT-US95-04570-2  
; Sequence 2, Application PC/TUS9504570  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Staskawicz, Brian J.  
; APPLICANT: Brent, Andrew F.  
; APPLICANT: Dahlbeck, Douglas  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kunkel, Barbara N.  
; APPLICANT: Mindrinos, Michael N.  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2904  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04570  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,360  
; FILING DATE: 13-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/230001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 100254  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04570-2

Query Match 50.0%; Score 49; DB 5; Length 885;  
Best Local Similarity 55.6%; Pred. No. 8.7;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19  
Db 103 ADYKLCCKVSAILKSIGE 120

RESULT 5  
PCT-US95-04589-2  
Sequence 2, Application PC/TUS9504589  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Staskawicz, Brian J.  
APPLICANT: Brent, Andrew F.  
APPLICANT: Dahlbeck, Douglas  
APPLICANT: Katagiri, Fumitaki  
APPLICANT: Kunkel, Barbara N.  
APPLICANT: Mindrinos, Michael N.  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: RPS2 GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 201  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street Suite 3100  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2904  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04589  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,360  
FILING DATE: 13-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/230001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 100254  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04589-2

Query Match 50.0%; Score 49; DB 5; Length 885;  
Best Local Similarity 55.6%; Pred. No. 8.7;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19  
Db 103 ADYKLCCKVSAILKSIGE 120

RESULT 6  
US-08-930-996A-7  
Sequence 7, Application US/08930996A  
Patent No. 6100449  
GENERAL INFORMATION:  
APPLICANT: FLUHR, Robert  
APPLICANT: ESHED, Yuval  
APPLICANT: ORI, Naomi  
APPLICANT: PARAN, Ilan  
APPLICANT: ZAMIR, Daniel  
TITLE OF INVENTION: A GENE FAMILY FROM THE I2 FUSARIUM RESISTANCE  
TITLE OF INVENTION: LOCUS OF TOMATO AND USE THEREOF FOR TRANSFORMATION AND  
TITLE OF INVENTION: SELECTIVE BREEDING OF TOMATO AND RELATED PLANTS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,996A  
FILING DATE: 09-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/05272  
FILING DATE: 15-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 113,373  
FILING DATE: 13-APR-1995  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-930-996A-7

Query Match 50.0%; Score 49; DB 3; Length 907;  
Best Local Similarity 55.6%; Pred. No. 8.9;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19  
Db 111 ADYKLCCKVSAILKSIGE 128

RESULT 7  
US-08-310-912A-142  
Sequence 142, Application US/08310912A  
Patent No. 5981730  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Staskawicz, Brian J.  
APPLICANT: Brent, Andrew F.  
APPLICANT: Dahlbeck, Douglas  
APPLICANT: Katagiri, Fumitaki  
APPLICANT: Kunkel, Barbara N.

```
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310,912A
; FILING DATE: September 22, 1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-310-912A-142

Query Match 50.0%; Score 49; DB 2; Length 909;
Best Local Similarity 55.6%; Pred. No. 8.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19
Db 111 ADYKLCCKVSAILKSIGE 128

RESULT 8
; Sequence 142, Application US/09301085
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
; FILE REFERENCE: 00786/254002
; CURRENT APPLICATION NUMBER: US/09/301,085
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 4.0

US-09-301-085-142
; Sequence 142, Application US/09301085
; Patent No. 6262248
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: DETECTION METHODS
; FILE REFERENCE: 00786/254002
; CURRENT APPLICATION NUMBER: US/09/301,085
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 4.0

US-09-301-085-142
; Sequence 142, Application US/09301085
; Patent No. 6262248
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND
; FILE REFERENCE: 00786/254002
; CURRENT APPLICATION NUMBER: US/09/301,085
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 4.0

US-09-301-085-142
; Sequence 142, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04589-142

Query Match 50.0%; Score 49; DB 5; Length 909;
Best Local Similarity 55.6%; Pred. No. 8.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19
Db 111 ADYKLCCKVSAILKSIGE 128

PCT-US95-04589-142
; Sequence 142, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04589-142

Query Match 50.0%; Score 49; DB 5; Length 909;
Best Local Similarity 55.6%; Pred. No. 8.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDGLGE 19
Db 111 ADYKLCCKVSAILKSIGE 128

PCT-US95-04589-142
```

```
RESULT 10
US-09-248-796A-16870
; Sequence 16870, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16870
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16870

Query Match      45.9%; Score 45; DB 4; Length 582;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDGL 18
   ||| : |||
Db 78 RDDYEFMIANPLTMKDLG 95

RESULT 11
US-09-328-352-5599
; Sequence 5599, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5599
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5599

Query Match      45.4%; Score 44.5; DB 4; Length 733;
Best Local Similarity 45.5%; Pred. No. 41;
Matches 10; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

Qy 1 RSDYKLYNKSST-----LKDL 17
   ||| : ||| : |||
Db 1 RQDFKLMNQNTTETIDLKEL 22

RESULT 12
US-09-248-796A-19901
; Sequence 19901, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16870
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16870

Query Match      43.9%; Score 43; DB 4; Length 820;
Best Local Similarity 53.3%; Pred. No. 84;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDGL 19
   ||| : |||
Db 559 KYNINQTVMKDLSE 573

RESULT 13
US-09-134-078-21
; Sequence 21, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134.078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-134-078-21

Query Match      42.9%; Score 42; DB 3; Length 511;
Best Local Similarity 41.2%; Pred. No. 71;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGL 18
   : ||| : |||
Db 58 NSVELYERDQEIADKL 74

RESULT 14
US-09-248-796A-14458
; Sequence 14458, Application US/09248796A
```

```
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14458
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14458

Query Match 42.9%; Score 42; DB 4; Length 591;
Best Local Similarity 46.7%; Pred. No. 84;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKDL 17
Db 442 DYQLFQKTNQNFKDL 456

RESULT 15
US-09-710-279-3262
; Sequence 3262, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3262
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3262

Query Match 42.3%; Score 41.5; DB 4; Length 176;
Best Local Similarity 37.5%; Pred. No. 25;
Matches 9; Conservative 3; Mismatches 7; Indels 5; Gaps 1;

Qy 1 RSDYKLYNKNS-----STLKDGLGE 19
Db 135 RNDYRFYPSNQEQRLVYKDLGQ 158

RESULT 16
US-09-134-001C-3379
; Sequence 3379, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; 
```

```
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3379
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3379

Query Match 42.3%; Score 41.5; DB 3; Length 357;
Best Local Similarity 37.5%; Pred. No. 57;
Matches 9; Conservative 3; Mismatches 7; Indels 5; Gaps 1;

Qy 1 RSDYKLYNKNS-----STLKDGLGE 19
Db 316 RNDYRFYPSNQEQRLVYKDLGQ 339

RESULT 17
US-09-809-665A-153
; Sequence 153, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Actinobacillus pleuropneumoniae
US-09-809-665A-153

Query Match 42.3%; Score 41.5; DB 4; Length 369;
Best Local Similarity 55.6%; Pred. No. 59;
Matches 10; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

Qy 1 RSDYKLY---NKNSSTLK 15
Db 139 RNDYKKGAGTNESTTK 156

RESULT 18
US-08-231-193A-11
; Sequence 11, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; FILE REFERENCE: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; 
```



; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231.193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1464 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-193A-11

Query Match 42.3%; Score 41.5; DB 2; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
||||:|||||  
Db 1184 YKLYSKH-FTLKXG 1197

RESULT 19  
US-08-486-273A-11  
; Sequence 11, Application US/08486273A  
; Patent No. 5985586  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTRO SUBUNITS, DNA  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486.273A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231.193  
; FILING DATE: 20-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062

; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1464 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-486-273A-11

Query Match 42.3%; Score 41.5; DB 2; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
||||:|||||  
Db 1184 YKLYSKH-FTLKXG 1197

RESULT 20  
US-08-940-086A-11  
; Sequence 11, Application US/08940086A  
; Patent No. 611091  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/940.086A  
; FILING DATE: 29-SEPT-97  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231.193  
; FILING DATE: 20-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9383C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 450-8499  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1464 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-940-086A-11

Query Match 42.3%; Score 41.5; DB 3; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
||||:|||||  
Db 1184 YKLYSKH-FTLKXG 1197

```

RESULT 21
US-08-436-332B-10
; Sequence 10, Application US/08436332B
; Patent No. 6130058
; GENERAL INFORMATION:
; APPLICANT: LE BOURDELLES, BEATRICE
; APPLICANT: MYERS, BEATRICE
; APPLICANT: WHITING, PAUL JOHN
; TITLE OF INVENTION: STABLY TRANSFECTED CELL LINE EXPRESSING
; TITLE OF INVENTION: NMDA RECEPTORS, AND NOVEL CLONED NMDA
; TITLE OF INVENTION: RECEPTOR SUBUNIT SEQUENCES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MICHAEL D. YABLONSKY - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVE., - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436.332B
; FILING DATE: 10-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: YABLONSKY, MICHAEL D.
; REGISTRATION NUMBER: 40,407
; REFERENCE/DOCKET NUMBER: T1210Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-4678
; TELEFAX: 732-594-4720
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1464 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-436-332B-10

Query Match 42.3%; Score 41.5; DB 3; Length 1464;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYNKSSTLKDLG 18
|||||.|||||
Db 1184 YKLYSKH-FTLKDKG 1197

RESULT 22
US-08-940-035A-11
; Sequence 11, Application US/08940035A
; Patent No. 6316611
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.

```

```

; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940.035A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1464 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-940-035A-11

Query Match 42.3%; Score 41.5; DB 3; Length 1464;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYNKSSTLKDLG 18
|||||.|||||
Db 1184 YKLYSKH-FTLKDKG 1197

RESULT 23
US-08-935-105A-11
; Sequence 11, Application US/08935105A
; Patent No. 6376660
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935.105A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

```

NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-935-105A-11

Query Match 42.3%; Score 41.5; DB 3; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1184 YKLYSKH-FTLKDKG 1197

RESULT 24  
US-09-648-797-11  
; Sequence 11, Application US/09648797  
; Patent No. 6469142  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; Ellis, Steven B.  
; Liaw, Chen W.  
; Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Heller Ehtman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/648,797  
; FILING DATE: 28-Aug-2000  
; PRIOR APPLICATION DATA:  
; FILING DATE: 29-SEPT-97  
; APPLICATION NUMBER: US 08/940,086A  
; FILING DATE: 20-APR-1994  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 20-APR-1994  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9383C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 450-8499  
; INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-09-648-797-11

Query Match 42.3%; Score 41.5; DB 4; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1184 YKLYSKH-FTLKDKG 1197

RESULT 25  
US-08-217-704C-2  
; Sequence 2, Application US/08217704C  
; Patent No. 6489124  
; GENERAL INFORMATION:  
; APPLICANT: Foldes, Robert  
; Fantaskie, Robert  
; Adams, Sally-Lyn  
; Kamboj, Rajender  
; TITLE OF INVENTION: MODULATORY PROTEINS OF HUMAN CNS  
RECEPTORS  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: Suite 500, 3000 K Street  
; CITY: Washington, D.C., N.W.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/217,704C  
; FILING DATE: 25-Mar-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/987,953  
; FILING DATE: 11-DEC-94  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16777/259/ALLE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-217-704C-2

Query Match 42.3%; Score 41.5; DB 4; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYKNSSTLKDLG 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1184 YKLYSKH-FTLKDKG 1197

RESULT 26  
US-09-386-123-11  
; Sequence 11, Application US/09386123  
; Patent No. 6521413  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun

;; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
;; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
;; NUMBER OF SEQUENCES: 63  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Heller Ehrman White & McAuliffe  
;; STREET: 4250 Executive Square, 7th Floor  
;; CITY: La Jolla  
;; STATE: CA  
;; COUNTRY: U.S.A.  
;; ZIP: 92037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/386,123  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/486,273  
;; FILING DATE: 06-JUNE-95  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/231,193  
;; FILING DATE: 20-APR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/052,449  
;; FILING DATE: 20-APR-1993  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6362-9383F  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 858-450-8403  
;; TELEFAX: 858-587-5360  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1464 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-386-123-11  
Query Match 42.3%; Score 41.5; DB 4; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
Qy 4 YKLYNKSSTLKDLG 18  
Db 1184 YKLYSKH-FTLKDG 1197  
RESULT 27  
US-09-270-767-58231  
; Sequence 58231, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 58231  
; LENGTH: 126  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-58231  
Query Match 41.8%; Score 41; DB 4; Length 126;

Best Local Similarity 64.3%; Pred. No. 20;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 5 KLYNKSSTLKDLG 18  
Db 79 KLYNKMSSSLKDLG 92  
RESULT 28  
US-09-634-238-370  
; Sequence 370, Application US/09634238  
; Patent No. 6544772  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka J.  
; APPLICANT: Bloksberg, Leonard, N.  
; APPLICANT: Lubbers, Mark W.  
; APPLICANT: Dekker, James  
; APPLICANT: Christenson, Anna C.  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul W.  
; APPLICANT: Reid, Julian R.  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Polynucleotides, materials incorporating  
; TITLE OF INVENTION: them and methods for using them.  
; FILE REFERENCE: 11000.1043U1  
; CURRENT APPLICATION NUMBER: US/09/634,238  
; CURRENT FILING DATE: 2000-08-08  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 370  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-370  
Query Match 41.8%; Score 41; DB 4; Length 182;  
Best Local Similarity 44.4%; Pred. No. 31;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
Qy 2 SDYKLYNKSSTLKDLGE 19  
Db 110 SDYQAINPTFTGWMKDFDE 127  
RESULT 29  
US-09-198-452A-1091  
; Sequence 1091, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffois, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 1091  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-1091  
Query Match 41.8%; Score 41; DB 4; Length 245;  
Best Local Similarity 50.0%; Pred. No. 44;  
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 RSDYKLYNKSSTLKDLG 18  
Db 1 RMSYFNQKNSVVLRLSLG 18

RESULT 30  
US-09-270-767-42905  
; Sequence 42905, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 42905  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-42905  
  
Query Match 41.8%; Score 41; DB 4; Length 273;  
Best Local Similarity 64.3%; Pred. No. 50;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 5 KLYNKNSSSTLKDLG 18  
|:|:|:|:|:|:|:  
Db 226 KLNEMMSSTLKDLG 239  
|:|:|:|:|:|:|:  
  
RESULT 31  
US-09-404-258-6  
; Sequence 6, Application US/09404258  
; Patent No. 6274353  
; GENERAL INFORMATION:  
; APPLICANT: YANG, Shuwei  
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR IMPROVED POLYNUCLEOTIDE  
; FILE REFERENCE: 089962/0102  
; CURRENT APPLICATION NUMBER: US/09/404,258  
; CURRENT FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
US-09-404-258-6  
  
Query Match 41.8%; Score 41; DB 3; Length 281;  
Best Local Similarity 61.5%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 5 KLYNKNSSSTLKDL 17  
|:|:|:|:|:|:|:  
Db 121 KYQKIESTLKDI 133  
|:|:|:|:|:|:|:  
  
RESULT 32  
US-09-853-701-6  
; Sequence 6, Application US/09853701  
; Patent No. 6699981  
; GENERAL INFORMATION:  
; APPLICANT: YANG, Shuwei  
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR IMPROVED POLYNUCLEOTIDE  
; FILE REFERENCE: 089962/0102  
; CURRENT APPLICATION NUMBER: US/09/853,701  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: US 09/404,258  
; PRIOR FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6

; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
US-09-853-701-6  
  
Query Match 41.8%; Score 41; DB 4; Length 281;  
Best Local Similarity 61.5%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 5 KLYNKNSSSTLKDL 17  
|:|:|:|:|:|:|:  
Db 121 KYQKIESTLKDI 133  
|:|:|:|:|:|:|:  
  
RESULT 33  
US-08-210-394-1  
; Sequence 1, Application US/08210394  
; Patent No. 5770213  
; GENERAL INFORMATION:  
; APPLICANT: Zlotnick Dr., Gary W.  
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus  
; TITLE OF INVENTION: Influenzae P5 Protein as a Vaccine for No. 5770213typable  
; TITLE OF INVENTION: Haemophilus Influenzae Strain  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: US  
; ZIP: 07470-8426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/210,394  
; APPLICATION NUMBER: US/08/210,394  
; FILING DATE: 07-MAR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J  
; REFERENCE/DOCKET NUMBER: 32,144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201/831-3246  
; TELEFAX: 201/831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-210-394-1  
  
Query Match 41.8%; Score 41; DB 1; Length 338;  
Best Local Similarity 61.5%; Pred. No. 65;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 RSDYKLYNKNSSST 13  
|:|:|:|:|:|:|:  
Db 119 RSDYKFYEAPNST 131  
|:|:|:|:|:|:|:  
  
RESULT 34  
US-09-543-681A-4728  
; Sequence 4728, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4728  
; LENGTH: 342  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4728

Query Match 41.8%; Score 41; DB 4; Length 342;  
Best Local Similarity 50.0%; Pred. No. 66;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDLG 18  
Db 126 RIHNKPSIEKDIG 139

RESULT 35  
US-09-198-452A-991  
; Sequence 991, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 991  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-991

Query Match 41.8%; Score 41; DB 4; Length 351;  
Best Local Similarity 61.5%; Pred. No. 68;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDL 17  
Db 322 KLYNQSKTTGKDL 334

RESULT 36  
US-09-248-796A-18781  
; Sequence 18781, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 18781  
; LENGTH: 388  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-18781

Query Match 41.8%; Score 41; DB 4; Length 388;  
Best Local Similarity 50.0%; Pred. No. 76;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKD 16  
Db 339 KSEYQSLKSGSTIKD 354

RESULT 37  
US-09-306-593-9  
; Sequence 9, Application US/09306593  
; Patent No. 6184018  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; APPLICANT: Ximenes, Eduardo A.  
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from  
; FILE REFERENCE: 31-98us  
; CURRENT APPLICATION NUMBER: US/09/306,593  
; CURRENT FILING DATE: 1999-05-06  
; EARLIER APPLICATION NUMBER: US 60/084,494  
; EARLIER FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patencin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Costus speciosus  
US-09-306-593-9

Query Match 41.8%; Score 41; DB 3; Length 476;  
Best Local Similarity 53.3%; Pred. No. 97;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKLYNKSSTLKDLG 18  
Db 63 YKXYKEDVKLLKDLG 77

RESULT 38  
US-09-134-078-18  
; Sequence 18, Application US/09134078  
; Patent No. 6368844  
; GENERAL INFORMATION:  
; APPLICANT: Bylina, Edward J.  
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
; STREET: 4365 Executive Drive, Suite 1600  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/134,078  
; FILING DATE: 13-AUG-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,026  
; FILING DATE: 10-OCT-1997  
; APPLICATION NUMBER: 60/056,916  
; FILING DATE: 06-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/024002  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 509 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-134-078-18

Query Match 41.8%; Score 41; DB 3; Length 509;  
Best Local Similarity 41.2%; Pred. No. 1e+02; 5; Indels 0; Gaps 0;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGL 18  
Db 58 NNYELYEKDRLARDLG 74

## RESULT 39

US-10-039-293A-2

; Sequence 2, Application US/10039293A

; Patent No. 6713281

; GENERAL INFORMATION:

; APPLICANT: SHORT, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC ENZYMES

; FILE REFERENCE: 112766-222 (DIV-1460-22)

; CURRENT APPLICATION NUMBER: US/10/039,293A

; CURRENT FILING DATE: 2001-12-31

; PRIOR APPLICATION NUMBER: US 09/185,373

; PRIOR FILING DATE: 1998-11-03

; PRIOR APPLICATION NUMBER: US 08/760,489

; PRIOR FILING DATE: 1996-12-05

; PRIOR APPLICATION NUMBER: US 60/008,311

; PRIOR FILING DATE: 1995-12-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 2

; LENGTH: 509

; TYPE: PRT

; ORGANISM: Thermococcus 9N2 Beta-glycosidase

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (215)..(215)

; OTHER INFORMATION: The 'Xaa' at location 215 stands for Glu, or Asp.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (224)..(224)

; OTHER INFORMATION: The 'Xaa' at location 224 stands for Gly.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (235)..(235)

; OTHER INFORMATION: The 'Xaa' at location 235 stands for Glu, Asp, or Ala.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (236)..(236)

; OTHER INFORMATION: The 'Xaa' at location 236 stands for Lys, or Asn.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (281)..(281)

; OTHER INFORMATION: The 'Xaa' at location 281 stands for Thr, Ala, Pro, or Ser.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (284)..(284)

; OTHER INFORMATION: The 'Xaa' at location 284 stands for Asn, Asp, His, or Tyr.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (294)..(294)

; OTHER INFORMATION: The 'Xaa' at location 294 stands for Thr, Ala, Pro, or Ser.

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (295)..(295)

; FEATURE:

OTHER INFORMATION: The 'Xaa' at location 295 stands for Thr, Ala, Pro, or Ser.  
US-10-039-293A-2

Query Match 41.8%; Score 41; DB 4; Length 509;  
Best Local Similarity 41.2%; Pred. No. 1e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDGL 18  
Db 58 NNYELYEKDRLARDLG 74

## RESULT 40

US-09-540-236-2373

; Sequence 2373, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 2373

; LENGTH: 579

; TYPE: PRT

; ORGANISM: M.catarrahals

; US-09-540-236-2373

Query Match 41.8%; Score 41; DB 4; Length 579;  
Best Local Similarity 53.3%; Pred. No. 1.2e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDGL 19  
Db 467 ELVKQNHGTLKDLGK 481

Search completed: November 24, 2004, 09:31:41

Job time : 17.7283 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:04 ; Search time 87.3587 Seconds  
(without alignments)  
125.141 Million cell updates/sec

Title: US-09-719-379A-4  
Perfect score: 98  
Sequence: 1 RSDYKLYNKNSSILKDLGE 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt.02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_crembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description          |
|------------|-------|---------------|--------|--------------|----------------------|
| 1          | 87    | 88.8          | 353    | 1 OM51_HAEIN | P43840 haemophilus   |
| 2          | 61    | 62.2          | 907    | 2 Q8LL02     | Q8L102 arabidopsis   |
| 3          | 50    | 51.0          | 289    | 2 Q8ZV60     | Q8ZV60 pyrobaculum   |
| 4          | 50    | 51.0          | 1043   | 2 Q7RLC4     | Q7RLC4 plasmodium    |
| 5          | 49    | 50.0          | 167    | 1 RPO7_FOWPV | Q05569 fowlpox vir   |
| 6          | 49    | 50.0          | 167    | 2 Q70GZ1     | Q70GZ1 fowlpox vir   |
| 7          | 49    | 50.0          | 167    | 2 CAE52706   | CAE52706 fowlpox v   |
| 8          | 49    | 50.0          | 179    | 2 Q7ZUH7     | Q7ZUH7 brachydanio   |
| 9          | 49    | 50.0          | 909    | 1 RPS2_ARATH | Q42484 arabidopsis   |
| 10         | 49    | 50.0          | 909    | 2 AAO67907   | AAO64907 arabidops   |
| 11         | 48    | 49.0          | 394    | 2 Q9QTH9     | Q9QTH9 svb2 plect    |
| 12         | 48    | 49.0          | 430    | 2 Q3YV75     | Q9YV75 melanoplus th |
| 13         | 48    | 49.0          | 450    | 2 Q6HMS5     | Q6HMS5 bacillus th   |
| 14         | 48    | 49.0          | 450    | 2 Q73C96     | Q73C96 bacillus ce   |
| 15         | 48    | 49.0          | 450    | 2 Q81U20     | Q81U20 bacillus an   |
| 16         | 48    | 49.0          | 450    | 2 AAS40100   | AAS40100 bacillus    |
| 17         | 48    | 49.0          | 450    | 2 AAS40173   | AAS40173 bacillus    |
| 18         | 48    | 49.0          | 505    | 1 SPKD_SYNY3 | P54735 synechocyst   |
| 19         | 48    | 49.0          | 549    | 2 Q8IK31     | Q8IK31 bacillus an   |
| 20         | 48    | 49.0          | 549    | 2 AAT34652   | AAT34652 bacillus    |
| 21         | 48    | 49.0          | 802    | 2 Q9QP04     | Q9QP04 human immun   |
| 22         | 48    | 49.0          | 851    | 2 Q9QB36     | Q9QB36 human immun   |
| 23         | 47    | 48.0          | 181    | 2 Q9V832     | Q9V832 drosophila    |
| 24         | 47    | 48.0          | 294    | 2 Q7ZMS2     | Q7ZMS2 xenopus lae   |
| 25         | 47    | 48.0          | 720    | 2 Q7MV25     | Q7MV25 porphyronon   |
| 26         | 46    | 46.9          | 268    | 2 Q6LX25     | Q6LX25 methanococ    |
| 27         | 46    | 46.9          | 268    | 2 CAF31084   | CAF31084 methanoco   |
| 28         | 46    | 46.9          | 450    | 2 Q81GW3     | Q81GW3 bacillus ce   |
| 29         | 46    | 46.9          | 465    | 2 Q9PJ81     | Q9PJ81 campylobact   |
| 30         | 46    | 46.9          | 530    | 2 Q8X0B9     | Q8X0B9 neurospora    |
| 31         | 46    | 46.9          | 784    | 2 Q8BJQ2     | Q8BJQ2 mus musculu   |

|    |      |      |      |   |            |                     |
|----|------|------|------|---|------------|---------------------|
| 32 | 46   | 46.9 | 784  | 2 | Q8VE17     | Q8VE17 mus musculu  |
| 33 | 46   | 46.9 | 784  | 2 | Q8VEM4     | Q8VEM4 mus musculu  |
| 34 | 46   | 46.9 | 1829 | 1 | DPOL_THEST | Q3845 thermococcu   |
| 35 | 46   | 46.9 | 2006 | 2 | Q7KSQ6     | Q7KSQ6 plasmodium   |
| 36 | 46   | 46.9 | 2006 | 2 | AAQ73469   | AAQ73469 plasmodium |
| 37 | 46   | 46.9 | 2019 | 2 | Q7K5Q5     | Q7K5Q5 plasmodium   |
| 38 | 46   | 46.9 | 2019 | 2 | AAQ73468   | AAQ73468 plasmodium |
| 39 | 46   | 46.9 | 2055 | 2 | Q8IHP3     | Q8IHP3 plasmodium   |
| 40 | 46   | 46.9 | 2055 | 2 | Q8TSC7     | Q8TSC7 plasmodium   |
| 41 | 46   | 46.9 | 4915 | 2 | Q6CJB6     | Q6CJB6 kluyveromyc  |
| 42 | 45.5 | 46.4 | 320  | 2 | Q6L205     | Q6L205 picrophilus  |
| 43 | 45   | 45.9 | 60   | 2 | Q72SX1     | Q72SX1 leptospira   |
| 44 | 45   | 45.9 | 60   | 2 | Q8F2K3     | Q8F2K3 leptospira   |
| 45 | 45   | 45.9 | 60   | 2 | AAS69857   | AAS69857 leptospir  |

#### ALIGNMENTS

##### RESULT 1

| ID | OM51_HAEIN   | STANDARD; | PRT; | 353 AA. |
|----|--|-----------|------|---------|
| AC | P43840;  |           |      |         |
| DT | 01-NOV-1995 (Rel. 32, Created)   |           |      |         |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update)  |           |      |         |
| DT | 01-OCT-2004 (Rel. 45, Last annotation update)  |           |      |         |
| DE | Outer membrane protein P5 precursor (OMP P5).  |           |      |         |
| GN | Name=ompA; Synonyms=ompP5; OrderedLocuNames=H1164;   |           |      |         |
| OS | Haemophilus influenzae.  |           |      |         |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;   |           |      |         |
| OC | Pasteurellaceae; Haemophilus.  |           |      |         |
| OX | NCBI_TaxID=727;  |           |      |         |
| RN | [1]  |           |      |         |
| RP | SEQUENCE FROM N.A.   |           |      |         |
| RC | STRAIN=Rd / KW20 / ATCC 51907;   |           |      |         |
| RC | MEDLINE=95350630; PubMed=7542800;  |           |      |         |
| RA | Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,   |           |      |         |
| RA | Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,   |           |      |         |
| RA | McKenny K., Sutton G.G., Fitzhugh W., Fields C.A., Gocayne J.D.,   |           |      |         |
| RA | Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,   |           |      |         |
| RA | Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  |           |      |         |
| RA | Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,   |           |      |         |
| RA | Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  |           |      |         |
| RA | Ghehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  |           |      |         |
| RA | Venter J.C.;   |           |      |         |
| RT | Rd."   |           |      |         |
| RL | Science 269:496-512(1995).   |           |      |         |
| CC | -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.   |           |      |         |
| CC | -1- SIMILARITY: Belongs to the ompA family.  |           |      |         |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration   |           |      |         |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |           |      |         |
| CC | the European Bioinformatics Institute. There are no restrictions on its  |           |      |         |
| CC | use by non-profit institutions as long as its content is in no way   |           |      |         |
| CC | modified and this statement is not removed. Usage by and for commercial  |           |      |         |
| CC | entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |           |      |         |
| CC | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |           |      |         |
| CC | EMBL: U32796; AAC22819.1; -  |           |      |         |
| DR | PIR: C64187; C64187.   |           |      |         |
| DR | HSSP: P02934; IQP.   |           |      |         |
| DR | TIGR: H1164; -   |           |      |         |
| DR | InterPro: IPR006664; Bac OmpA.   |           |      |         |
| DR | InterPro: IPR002368; OmpA.   |           |      |         |
| DR | InterPro: IPR006665; OmpA/MotB.  |           |      |         |
| DR | InterPro: IPR006690; OmpA_LIKE.  |           |      |         |
| DR | InterPro: IPR000498; OmpA_tmam.  |           |      |         |
| DR | Pfam: PF00691; OmpA; 1.  |           |      |         |
| DR | PRINTS: PR01021; OMPADOMAIN.   |           |      |         |
| DR | PRODOM: PD000930; OmpA/MotB; 1.  |           |      |         |
| DR | PROSITE: PS01068; OmpA; 1.   |           |      |         |

```

KW Complete proteome; Outer membrane; Porin; Signal; Transmembrane.
FT SIGNAL 1 21 By similarity.
FT CHAIN 22 353 Outer membrane protein P5.
FT DISULFID 326 338 By similarity.
FT DOMAIN 272 316 OmpA-like.
SQ SEQUENCE 353 AA; 37743 MW; 6ACB3E7BFF96B39 CRC64;

Query Match 88.8%; Score 87; DB 1; Length 353;
Best Local Similarity 89.5%; Pred. No. 3.9e-05;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19
    |||||:|||||
Db 132 RSDYKLYNKSSTLKDLGE 150

RESULT 2
QBLL02 PRELIMINARY; PRT; 907 AA.
AC QBLL02;
AC QBLL02;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE RPS2;
GN Name=rps2;
OS Arabidopsis lyrata (lyre-leaved rock-cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=59689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22505406; PubMed=12618410;
RA Mauricio R., Stahl E.A., Korves T., Tian D., Kreitman M.,
RA Bergelson J.;
RT "Natural selection for polymorphism in the disease resistance gene
RT Rps2 of Arabidopsis thaliana.";
RL Genetics 163:735-746 (2003).
DR EMBL; AF487796; AAM90858.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:000166; F:nucleotide binding; IEA.
DR GO; GO:0006935; P:apoptosis; IEA.
DR GO; GO:0042829; P:defense response to pathogen; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR00767; Disease_resist.
DR InterPro; IPR01611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 8.
DR Pfam; PF00931; NB-ARC; 1.
DR PRINTS; PR00364; DISEASERSIST.
DR SMART; SM00382; AAA; 1.
KW ATP-binding.
SQ SEQUENCE 907 AA; 104220 MW; EB3147059171FCC3 CRC64;

Query Match 62.2%; Score 61; DB 2; Length 907;
Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDLGE 19
    :|||||:|||||
Db 111 ADYKLCNKVSTLKSIGE 128

RESULT 3
QBZV60 PRELIMINARY; PRT; 289 AA.
AC QBZV60;
AC QBZV60;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ornithine cyclodeaminase (ArCB).
GN OrderedLocustNames=PAE2445;
OS Pyrobaculum aerophilum.

```

```

OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989 (2002).
DR EMBL; AE009876; AAL64196.1; -.
DR InterPro; IPR003462; ODC_Mu_crystatall.
DR Pfam; PF02423; ODC_Mu_crystatall; 1.
KW Complete proteome.
SQ SEQUENCE 289 AA; 30809 MW; 4914994B4FD110C6 CRC64;

Query Match 51.0%; Score 50; DB 2; Length 289;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YKLYNKSSTLKDLGE 19
    |||||:|||||
Db 141 YKLYNKSSTLKDLGE 156

RESULT 4
Q7RLC4 PRELIMINARY; PRT; 1043 AA.
ID Q7RLC4;
AC Q7RLC4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein (Fragment).
GN Name=PY02622;
GN Plasmodium yoelii yoelii.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=17XNL;
RX PubMed=12368665;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC ENBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000722; EAA22094.1; -.
DR Hypothetical protein.
KW NON TER 1043 1043
SQ SEQUENCE 1043 AA; 122316 MW; 0CE7D5C185132E2E CRC64;

Query Match 51.0%; Score 50; DB 2; Length 1043;
Best Local Similarity 55.6%; Pred. No. 77;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDLGE 19
    :|||||:|||||
Db 763 NDYLENENKSTKEKIGE 780

```

```

RESULT 5
RP07_FOPV STANDARD; PRT; 167 AA.
AC Q05569; O93019;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA-directed RNA polymerase 19 kDa polypeptide (EC 2.7.7.6).
GN Name=RP019; Synonym=FPV169, FPA5, FP-A5;
OS Fowlpox virus (FPV)
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90209363; PubMed=2157318;
RA Kumar S., Boyle D.B.;
RT "Mapping of a major early/late gene of fowlpox virus.";
RL Virus Res. 15:175-186(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831(2000).
RN [3]
RP SEQUENCE OF 1-97 FROM N.A.
RC STRAIN=FP-9 / Isolate HP-440;
RX MEDLINE=98080402; PubMed=9420213;
RA Boulanger D., Green P., Smith T., Czerny C.P., Skinner M.A.;
RT "The 131-amino-acid repeat region of the essential 39-kilodalton core protein of fowlpox virus FP9, equivalent to vaccinia virus A4L protein, is nonessential and highly immunogenic.";
RL J. Virol. 72:170-179(1998).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
CC -!- SUBUNIT: This enzyme consists of at least eight subunits.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X52461; CAA36694.1; -.
DR EMBL; AF198100; AAF44513.1; -.
DR EMBL; AJ005164; CAA06407.1; -.
DR PIR; A60013; A60013.
DR InterPro; IPR007984; Pox_RNA_Pol_19.
DR Pfam; PF05320; Pox_RNA_Pol_19; 1.
DR PIRSF; PIRSF000743; RPO19; 1.
KW DNA-directed RNA polymerase; Early protein; Transcription; Transference.
SQ SEQUENCE 167 AA; 19500 MW; D67E10A724179C37 CRC64;

Query Match 50.0%; Score 49; DB 1; Length 167;
Best Local Similarity 38.9%; Pred. No. 17;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSTLKDGE 19
Db 16 NDYKSYDEDDSDISDGE 33

RESULT 6
Q70GZ1 PRELIMINARY; PRT; 167 AA.
ID Q70GZ1
AC Q70GZ1

```

```

DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE ASR RPO19 orthologue.
GN Name=fp9.169;
OS Fowlpox virus (isolate HP-438[Munich]).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10263;
RN [1]
RP SEQUENCE FROM N.A.
RX Skinner M.A., Laidlaw S.M.;
RT "Comparison of the genome sequence of FP9, an attenuated, tissue culture-adapted European fowlpox virus, with those of virulent American and European viruses.";
RL J. Gen. Virol. 85:305-322(2004).
DR EMBL; AJ581527; CAE52706.1; -.
DR InterPro; IPR007984; Pox_RNA_Pol_19.
DR Pfam; PF05320; Pox_RNA_Pol_19; 1.
DR PIRSF; PIRSF000743; RPO19; 1.
DR PRODOM; PD012915; Pox_RNA_Pol_19; 1.
SQ SEQUENCE 167 AA; 19500 MW; D67E10A724179C37 CRC64;

Query Match 50.0%; Score 49; DB 2; Length 167;
Best Local Similarity 38.9%; Pred. No. 17;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSTLKDGE 19
Db 16 NDYKSYDEDDSDISDGE 33

RESULT 7
CAE52706 PRELIMINARY; PRT; 167 AA.
ID CAE52706
AC CAE52706;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE ASR RPO19 orthologue.
GN FP9.169.
OS Fowlpox virus (isolate HP-438[Munich]).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus; Fowlpox virus.
OX NCBI_TaxID=10263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPI-438 Munich;
RA Skinner M.A., Laidlaw S.M.;
RT "Comparison of the genome sequence of FP9, an attenuated, tissue culture-adapted European fowlpox virus, with those of virulent American and European viruses.";
RL J. Gen. Virol. 85:305-322(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HPI-438 Munich;
RA Skinner M.A.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ581527; CAE52706.1; -.
KW Serine protease inhibitor.
SQ SEQUENCE 167 AA; 19500 MW; D67E10A724179C37 CRC64;

Query Match 50.0%; Score 49; DB 2; Length 167;
Best Local Similarity 38.9%; Pred. No. 17;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSTLKDGE 19
Db 16 NDYKSYDEDDSDISDGE 33

RESULT 8
Q7ZUH7

```

ID Q7ZUH7 PRELIMINARY; PRT; 179 AA.  
AC Q7ZUH7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein zgc:56526.  
GN Name=zgc:56526;  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=whole body;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Alechul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson K.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska J., Smalish D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=whole body;  
RA Strausberg R.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC049039; AAH49039.1; -;  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR InterPro; IPR002942; S4.  
DR Pfam; PF01479; S4; 1.  
DR SMART; SM00363; S4; 1.  
DR PROSITE; PS50889; S4; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 179 AA; 20996 MW; 9960F6A2EB5BA3AF CRC64;

Query Match 50.0%; Score 49; DB 2; Length 179;  
Best Local Similarity 47.4%; Pred. No. 18;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RSDYKLYKNKSTLKDLGE 19  
||| ||| ||| :  
Db 41 RSDYKLYKNKSRNRDLAQ 59

RESULT 9  
RP52 ARATH  
ID RP52 ARATH STANDARD; PRT; 909 AA.  
AC Q42484; Q82096; Q8L3R0; Q8L3W3; Q8L4X9; Q8L4Y0; Q8L587; Q8L5B3;  
AC Q8LKZ8; Q8LKZ9; Q8LL00; Q8LL01; Q9ASP5;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE Disease resistance protein RP52 (Resistance to Pseudomonas syringae  
protein 2).  
GN Name=RP52; OrderedLocuNames=At4g26090; ORFNames=F20B18.200;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=94377978; PubMed=8091210;  
RA Bent A.F., Kunkel B.N., Dahlbeck D., Brown K.L., Schmidt R.,  
RA Giraudat J., Leung J., Staskawicz B.J.;  
RT "RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant  
disease resistance genes.";  
RL Science 265:1856-1860 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=95007758; PubMed=7923358;  
RA Mindrinos M., Katagiri F., Yu G.-L., Ausubel F.M.;  
RT "The A. thaliana disease resistance gene RPS2 encodes a protein  
containing a nucleotide-binding site and leucine-rich repeats.";  
RL Cell 78:1089-1099 (1994).  
RN [3]  
RP SEQUENCE FROM N.A., AND VARIANTS.  
RC STRAIN=cv. Po-1;  
RX MEDLINE=21231631; PubMed=11333251;  
RA Banerjee D., Zhang X., Bent A.F.;  
RT "The leucine-rich repeat domain can determine effective interaction  
between RPS2 and other host factors in Arabidopsis RPS2-mediated  
disease resistance.";  
RL Genetics 158:439-450 (2001).  
RN [4]  
RP SEQUENCE FROM N.A., AND VARIANTS.  
RC STRAIN=cv. Ab-7, cv. Ang-0, cv. Bg-4, cv. Bla-2, cv. Bur-0, cv. C2-1,  
RC cv. Co-1, cv. Ct-1, cv. Cvi-0, cv. D2-9, cv. Fm-17, cv. G2-1,  
RC cv. Golt-20, cv. Gs-6, cv. Hs-12, cv. Kas-1, cv. KNO2, cv. Mt-0,  
RC cv. Tamm-17, cv. Tsu-0, cv. Po-1, cv. Pog-0, cv. Pu-8, cv. RLD,  
RC cv. Yo-0, cv. Wu-0, and cv. Zu-0;  
RX MEDLINE=22505406; PubMed=12618410;  
RA Mauricio R., Stahl E.A., Korves T., Tian D., Kreitman M.,  
RA Bergelson J.;  
RT "Natural selection for polymorphism in the disease resistance gene  
rps2 of Arabidopsis thaliana.";  
RL Genetics 163:735-746 (2003).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083488; PubMed=10617198;  
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,  
RA Weichelgartner M., de Simone V., Obermaier B., Mache R., Schmidt T.,  
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,  
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
RA Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbusche F.,  
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbusche F.,  
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
RA Moolenaar P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
RA Berner S., Hempel S., Feldpausch M., Lamberth S., van den Daele H.,  
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., de Clercq R.,  
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Farmann B., Grandrath K., Dauner D., Herli A.,  
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
RA Massenon O., Argirou F., Clabaud G., Muehlen A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,  
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bieleke C.,





```

Db      85 EYKLYNRQDSTIVD 98
      :|||||: |||: |
      85 EYKLYNRQDSTIVD 98

RESULT 15
Q81U20      PRELIMINARY;          PRT;      450 AA.
AC  Q81U20; Q6I2A5; Q6KW36;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE  Hypothetical protein.
GN  OrderedLocusNames=BA1073; BAS1002; ORFNames=GBAA1073;
OS  Bacillus anthracis.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=1392;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=Ames / isolate Porton;
RX  MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA  Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA  Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA  Holtzapple E.K., Okstad O.A., Helgason E., Rikstone J., Wu M.,
RA  Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA  DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA  Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA  Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA  Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA  Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA  Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,
RA  Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT  "The genome sequence of Bacillus anthracis Ames and comparison to
RT  closely related bacteria."
RL  Nature 423:81-86 (2003).
RN  (2)
RP  SEQUENCE FROM N.A.
RC  STRAIN=Ames / isolate 0581;
RA  Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA  Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA  Fraser C.M.;
RT  "Bacillus anthracis comparative genomics."
RL  Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN  (3)
RP  SEQUENCE FROM N.A.
RC  STRAIN=Sterne;
RA  Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA  Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA  Richardson P., Rubin E., Tice H.;
RL  Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE017027; AAP25053.1; -.
DR  EMBL; AE017334; AAT30173.1; -.
DR  EMBL; AE017225; AAT53326.1; -.
DR  TIGR; BA1073; -.
KW  Hypothetical protein.
SQ  SEQUENCE 450 AA; 52228 MW; E74D611E23BD9B62 CRC64;

Query Match      49.0%; Score 48; DB 2; Length 450;
Best Local Similarity 57.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      3 DYKLYNKSSTLKD 16
      :|||||: |||: |
      85 EYKLYNRQDSTIVD 98

RESULT 16
AAS40100      PRELIMINARY;          PRT;      450 AA.
AC  AAS40100;
DT  02-MAR-2004 (TrEMBLrel. 27, Created)
DT  02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT  04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE  Hypothetical protein.

```

```

GN  BCE1170.
OS  Bacillus cereus (strain ATCC 10987).
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=222523;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  PubMed=14960714;
RA  Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA  Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA  Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT  "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT  adaptations and a large plasmid related to Bacillus anthracis pXOI."
RL  Nucleic Acids Res. 32:977-988 (2004).
DR  EMBL; AE017268; AAS40100.1; -.
DR  TIGR; BCE1170; -.
KW  Hypothetical protein.
SQ  SEQUENCE 450 AA; 52152 MW; E7E33677CFC5C88C CRC64;

Query Match      49.0%; Score 48; DB 2; Length 450;
Best Local Similarity 57.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      3 DYKLYNKSSTLKD 16
      :|||||: |||: |
      85 EYKLYNRQDSTIVD 98

Db      85 EYKLYNRQDSTIVD 98

RESULT 17
AAT30173      PRELIMINARY;          PRT;      450 AA.
AC  AAT30173;
DT  01-JUN-2004 (TrEMBLrel. 27, Created)
DT  01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT  01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE  Hypothetical protein.
GN  GBAA1073.
OS  Bacillus anthracis str. Ames 0581.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OX  NCBI_TaxID=261594;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=Ames 0581;
RA  Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA  Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT  "Bacillus anthracis comparative genomics."
RL  Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE017334; AAT30173.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 450 AA; 52228 MW; E74D611E23BD9B62 CRC64;

Query Match      49.0%; Score 48; DB 2; Length 450;
Best Local Similarity 57.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      3 DYKLYNKSSTLKD 16
      :|||||: |||: |
      85 EYKLYNRQDSTIVD 98

Db      85 EYKLYNRQDSTIVD 98

RESULT 18
SPKD_SINY3      STANDARD;          PRT;      505 AA.
AC  P54735;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  05-JUL-2004 (Rel. 44, Last annotation update)
DE  Serine/threonine-protein kinase D (EC 2.7.1.37).
GN  Name=spkD; OrderedLocusNames=sll10776;
OS  Synechocystis sp. (strain PCC 6803).
OC  Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX  NCBI_TaxID=1148;
RN  (1)

```

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=22158630; PubMed=12168951;  
 RA Kamei A., Yusea T., Geng X., Ikeuchi M.;  
 RT "Biochemical examination of the potential eukaryotic-type protein  
 RT kinase genes in the complete genome of the unicellular Cyanobacterium  
 RT *synechocystis* sp. PCC 6803.";  
 RL DNA Res. 9:71-78(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- SIMILARITY: Belongs to the Ser/thr protein kinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB046600; BAB17036.1; -;  
 DR EMBL; D64005; BAB10726.1; -;  
 DR F01; S77034; S77034.  
 DR InterPro; IPR011009; Kinase like.  
 DR InterPro; IPR000719; Prot Kinase.  
 DR InterPro; IPR008271; Ser thr\_pkin\_AS.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; P000001; Prot\_kinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Complete proteome; Serine/threonine-protein kinase;  
 KW Transferase.  
 FT DOMAIN 9 271 Protein kinase.  
 FT NP\_BIND 15 23 ATP (By similarity).  
 FT BINDING 40 40 ATP (By similarity).  
 FT ACT\_SITE 136 136 By similarity.  
 SQ SEQUENCE 505 AA; 55213 MW; C4F12A1896C4D51C CRC64;  
 Query Match 49.0%; Score 48; DB 1; Length 505;  
 Best Local Similarity 60.0%; Pred. No. 75;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 KLYNKNSSSTLKLGE 19  
 DB 58 KLPEKEASVLEDLGE 72  
 RESULT 19  
 Q81K31  
 ID Q81K31 PRELIMINARY; PRT; 549 AA.  
 AC Q81K31; Q6HQN1; Q6KK04;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Teichoic acid ABC transporter, ATP-binding protein.  
 GN OrderedLocNames=BA5510, BA55118, ORFNAMES=GBAA5510;  
 OS Bacillus anthracis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 EX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01596;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillie D.W., Faulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,

RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
 RA Kolony J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Plamondon Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
 RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,  
 RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics.";  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sterne; Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family.  
 DR EMBL; AE017041; AAP29155.1; -;  
 DR EMBL; AE017344; AAP24652.1; -;  
 DR EMBL; AE017225; AAT57407.1; -;  
 DR TIGR; BA5510; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; P:ATP binding; IEA.  
 DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO; GO:0000166; P:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR001452; SH3\_bac.  
 DR InterPro; IPR003646; SH3\_bac.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR SMART; SM00287; SH3b; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; UNKNOWN\_1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 549 AA; 61970 MW; 864A93BCEBBEB0DB CRC64;  
 Query Match 49.0%; Score 48; DB 2; Length 549;  
 Best Local Similarity 71.4%; Pred. No. 82;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 YKLYNKNSSSTLKL 17  
 DB 14 YKLYNKPFDKLKL 27  
 RESULT 20  
 AAT34652  
 ID AAT34652 PRELIMINARY; PRT; 549 AA.  
 AC AAT34652;  
 DT 01-JUN-2004 (TrEMBLrel. 27, Created)  
 DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Teichoic acid ABC transporter, ATP-binding protein.  
 GN GBAA5510.  
 OS Bacillus anthracis str. Ames 0581.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
 OC Bacillus cereus group; Bacillus anthracis.  
 OX NCBI\_TaxID=261594;  
 RN [1]



```

RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasako D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RW EMBL; AE017334; AAT34652.1; -.
KW ATP-binding.
SQ SEQUENCE 549 AA; 61970 MW; 864A93BCBBEBB0DB CRC64;

Query Match 49.0%; Score 48; DB 2; Length 549;
Best Local Similarity 71.4%; Pred. No. 82;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 YKLYNKNSSTLKDL 17
Db 14 YKLYNKPPDKDL 27

RESULT 21
Oy Q9QP04 PRELIMINARY; PRT; 802 AA.
AC Q9QP04;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Envelope (Gp160) (Fragment).
GN Name=gp160;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MP411;
RA Triques K., Bourgeois A., Saragosti S., Vidal N., Mpoudi-Ngole E.,
RA Nzilambi N., Apetrei C., Ekwilanga M., Delaporte E., Peeters M.;
RT "High diversity of HIV-1 subtype F strains in Central Africa.";
RL Virology 259:99-109(1999).
DR EMBL; AJ237804; CAB44055.2; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT NON TER 1
FT CHAIN 802
SQ SEQUENCE 802 AA; 90885 MW; DB27A3076DACF194 CRC64;

Query Match 49.0%; Score 48; DB 2; Length 802;
Best Local Similarity 64.3%; Pred. No. 1.2e+02;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 2 SDYKLYNKNSSTLK 15
Db 159 SDYRLINCNTSTIK 172

RESULT 22
Oy Q9QBY6 PRELIMINARY; PRT; 851 AA.
AC Q9QBY6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Envelope (Gp160) (Fragment).
GN Name=env;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

```

```

OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20192179; PubMed=10659053;
RA Triques K., Bourgeois A., Vidale N., Mpoudi-Ngole E.,
RA Mulanga-Kabeya C., Nzilambi N., Torimiro N., Saman E., Delaporte E.,
RA Peeters M.;
RT "Near-full-length genome sequencing of divergent African HIV type 1
RT subtype F viruses leads to the identification of a new HIV type 1
RT subtype designated K.";
RL AIDS Res. Hum. Retroviruses 16:139-151(2000).
DR EMBL; AJ249238; CAB58986.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT NON TER 1
FT CHAIN 847 851 NEF protein.
FT CHAIN <1 851 NEF protein.
SQ SEQUENCE 851 AA; 96574 MW; 4B091F2348930F4C CRC64;

Query Match 49.0%; Score 48; DB 2; Length 851;
Best Local Similarity 64.3%; Pred. No. 1.3e+02;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 2 SDYKLYNKNSSTLK 15
Db 198 SDYRLINCNTSTIK 211

RESULT 23
Oy Q9V832 PRELIMINARY; PRT; 181 AA.
AC Q9V832; Q8SVH9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CG4866-PA (RE57564p).
GN ORFNames=CG4866;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoeltin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.M., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RT Science 287:2185-2195(2000).  
RN [2]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT melanogaster euchromatic genome sequence.";  
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RN [4]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RN [5]  
RN  
RP SEQUENCE FROM N.A.  
RX FLYBASE;  
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RN  
RP SEQUENCE FROM N.A.  
RX FLYBASE;  
RN Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dreesnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003803; AAF57847.2; -;  
DR EMBL; AY01534; AAL49156.1; -;  
DR InAct; Q9V832; -;  
DR FLYBase; FBgn0034232; CG4866.

DR GO; GO:0003723; F:RNA binding; IEA.  
DR InterPro; IPR002942; S4.  
DR Pfam; PF01479; S4; 1.  
DR SMART; SM00363; S4; 1.  
DR PROSITE; PSS00889; S4; 1.  
SQ SEQUENCE 181 AA; 21304 MW; 1EAD137DEC9A54AA CRC64;  
  
Query Match 48.0%; Score 47; DB 2; Length 181;  
Best Local Similarity 47.4%; Pred. No. 37;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 1 RSDYKLYNKSSTLKDLGE 19  
Db 41 REDYTKYNKLSREIRELAE 59  
  
RESULT 24  
Q7ZWS2 PRELIMINARY; PRT; 294 AA.  
AC Q7ZWS2;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DS LOC398553 protein (Fragment).  
GN Name=LOC398553;  
OS *Xenopus laevis* (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; *Xenopus*.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RN  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=22341132; PubMed=12454917;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus*  
RT initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
RN [3]  
RN  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RA Klein S., Strausberg R.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC046729; AAH46729.1; -;  
DR InterPro; IPR006434; HAD\_SF-IE.  
DR Pfam; PF05822; UMPH-1; 1\_  
DR TIGRFAMs; TIGR01544; HAD-SF-IE; 1.  
FT NON\_TER 1 1

```

SQ SEQUENCE 294 AA; 33984 MW; 7291143809896760 CRC64;
Query Match 48.0%; Score 47; DB 2; Length 294;
Best Local Similarity 90.0%; Pred. No. 61;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 YKNKSSTLKD 16
Db 205 YKNSSVLKD 214

RESULT 25
Q7MV25 PRELIMINARY; PRT; 720 AA.
AC Q7MV25;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=PG1283;
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=22829867; PubMed=12949112;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.L.,
RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T.M., Tallon L., Gray J.,
RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA Dewhirst F.E., Fraser C.M.;
RT "Complete genome sequence of the oral pathogenic bacterium
RT Porphyromonas gingivalis strain W83."
RL J. Bacteriol. 185:5591-5601(2003).
DR EMBL; AE017176; AAQ66363.1; -.
DR TIGR; PG1283; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 720 AA; 81982 MW; D0FF2762BBE0901D CRC64;

Query Match 48.0%; Score 47; DB 2; Length 720;
Best Local Similarity 57.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKD 16
Db 47 EYDLYNPGTSLKD 60

RESULT 26
Q6LX25 PRELIMINARY; PRT; 268 AA.
AC Q6LX25;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Prephenate dehydratase (EC 4.2.1.51).
GN Name=pheA; OrderedLocusNames=MPI1528;
OS Methanococcus maripaludis
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / LL;
RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeiri A., Rouse G.,
RA Saenphimmachak C., Soll D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957223; CAF31084.1; -.
KW Lyase.
SQ SEQUENCE 268 AA; 30314 MW; 44034DD35B0E1155 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 268;
Best Local Similarity 53.3%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKD 17
Db 238 DYEYDDNSALLRDL 252

RESULT 27
CAF31084 PRELIMINARY; PRT; 268 AA.
AC CAF31084;
DT 03-MAR-2004 (TrEMBLrel. 27, Created)
DT 03-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Prephenate dehydratase (EC 4.2.1.51).
GN PHEA OR MPI1528.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / LL;
RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeiri A., Rouse G.,
RA Saenphimmachak C., Soll D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957223; CAF31084.1; -.
KW Lyase.
SQ SEQUENCE 268 AA; 30314 MW; 44034DD35B0E1155 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 268;
Best Local Similarity 53.3%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKD 17
Db 238 DYEYDDNSALLRDL 252

RESULT 28
Q81GW3 PRELIMINARY; PRT; 450 AA.
AC Q81GW3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BC1071;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

RT methanogen Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957223; CAF31084.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0004664; F:prephenate dehydratase activity; IEA.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR008237; PDT.ACT.
DR InterPro; IPR001086; Preph_dehydratase.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00800; PDT; 1.
DR PROSITE; PROSITE01424; PDT.ACT; 1.
DR PROSITE; PROSITE00858; PREPHENATE_DEHYDR_2; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 268 AA; 30314 MW; 44034DD35B0E1155 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 268;
Best Local Similarity 53.3%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKD 17
Db 238 DYEYDDNSALLRDL 252

RESULT 27
CAF31084 PRELIMINARY; PRT; 268 AA.
AC CAF31084;
DT 03-MAR-2004 (TrEMBLrel. 27, Created)
DT 03-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Prephenate dehydratase (EC 4.2.1.51).
GN PHEA OR MPI1528.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / LL;
RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeiri A., Rouse G.,
RA Saenphimmachak C., Soll D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957223; CAF31084.1; -.
KW Lyase.
SQ SEQUENCE 268 AA; 30314 MW; 44034DD35B0E1155 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 268;
Best Local Similarity 53.3%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKD 17
Db 238 DYEYDDNSALLRDL 252

RESULT 28
Q81GW3 PRELIMINARY; PRT; 450 AA.
AC Q81GW3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BC1071;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.C.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AE017001; AAP08056.1; -.
SQ SEQUENCE 450 AA; 52193 MW; B93F7BE8FF57545 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 450;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSSTLK 16
Db 85 EYKYNRQDSIVD 98

RESULT 29
Q9PJ81 PRELIMINARY; PRT; 465 AA.
AC Q9PJ81.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Cj0030.
GN OrderedLocustNames=Cj0030;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Ketley J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagals K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139074; CAB72523.1; -.
DR PIR; A81419; A81419.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 465 AA; 56341 MW; E14E40D639E8D5C1 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 465;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 7 YNKNSSTLKDLGE 19
Db 271 YNKNSSELKLNLE 283

RESULT 30
Q8X0E9 PRELIMINARY; PRT; 530 AA.
AC Q8X0E9.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein B14A6.080.
GN Name=B14A6.080;

```

```

OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Partmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670007; CAD21310.1; -.
DR InterPro; IPR001810; F-box.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 530 AA; 59352 MW; 70382EE15F71BB9D CRC64;

Query Match 46.9%; Score 46; DB 2; Length 530;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 YKLYNKSSSTLK 15
Db 358 YRLFNKFSSTLK 369

RESULT 31
Q8BJQ2 PRELIMINARY; PRT; 784 AA.
AC Q8BJQ2.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 4 days neonate male adipose cDNA, RIKEN full-length
DE enriched library, clone:B430116N19 product:UBIQUITIN-SPECIFIC PROTEASE
DE homolog.
GN Name=Usp1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Adipose;
RX MEDLINE=92279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Adipose;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Adipose;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Adipose;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";

```

```

RC  Genome Res. 10:1617-1630(2000).
RN  [5]
RP  SEQUENCE FROM N.A.
RX  STRAIN=C57BL/6J; TISSUE=Adipose;
RC  MEDLINE=20530913; PubMed=11076861;
RA  Shiba K., Itoh M., Aizawa K., Nishida K., Kiteunai T., Tashiro H., Itoh M.,
RA  Konno H., Akiyama J., Aizawa K., Nishida K., Hazama M., Nishine T., Harada A.,
RA  Sumi N., Ishii Y., Nakamura S., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA  Yamamoto R., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA  Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA  Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayaishizaki Y.,
RT  "RIKEN integrated sequence analysis (RISA) system-384-format
RT  sequencing pipeline with 384 multiplexed sequencer.";
RL  Genome Res. 10:1757-1771(2000).
RN  [6]
RP  SEQUENCE FROM N.A.
RX  STRAIN=C57BL/6J; TISSUE=Adipose;
RC  STRAIN=C57BL/6J; TISSUE=Adipose;
RA  Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA  Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA  Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA  Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA  Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA  Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA  Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA  Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA  Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA  Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA  Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayaishizaki Y.;
RL  Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AK080882; BAC38059.1; --
DR  HSP; Q93009; INBF.
DR  MG; MG12385198; Uspl.
DR  GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR  GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.
DR  GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.
DR  GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR  InterPro; IPR000345; CytC_heme_BS.
DR  InterPro; IPR001394; Peptidase_C19.
DR  Pfam; PF00443; UCH_1.
DR  PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR  PROSITE; PS00972; UCH_2_1; 1.
DR  PROSITE; PS00973; UCH_2_2; 1.
DR  PROSITE; PS0235; UCH_2_3; 1.
KW  Protease.
SQ  SEQUENCE 784 AA; 87455 MW; 56F849DC1ADE3D81 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 784;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy  1 RSDYKLYNKNSSTLKDLG 18
Db  :||:|||||:|:|:|
    673 KADYELYNKASNPDKVVG 690

RESULT 32
Q8VE17 PRELIMINARY; PRT; 784 AA.
AC Q8VE17
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Uspl protein.
GN Name=Uspl.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
virgin mouse. Taken by biopsy.;
MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=FVB/N;
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
RC virgin mouse. Taken by biopsy.;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC020007; AAH20007.1; --
DR  HSP; Q93009; INBF.
DR  MG; MG12385198; Uspl.
DR  GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR  GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.
DR  GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR  InterPro; IPR000345; CytC_heme_BS.
DR  InterPro; IPR001394; Peptidase_C19.
DR  Pfam; PF00443; UCH_1.
DR  PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR  PROSITE; PS00972; UCH_2_1; 1.
DR  PROSITE; PS00973; UCH_2_2; 1.
DR  PROSITE; PS0235; UCH_2_3; 1.
SQ  SEQUENCE 784 AA; 87459 MW; 09656ACA2E36943B CRC64;

Query Match 46.9%; Score 46; DB 2; Length 784;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy  1 RSDYKLYNKNSSTLKDLG 18
Db  :||:|||||:|:|:|
    673 KADYELYNKASNPDKVVG 690

RESULT 33
Q8VEW4 PRELIMINARY; PRT; 784 AA.
AC Q8VEW4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ubiquitin specific protease 1.
GN Name=Uspl.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
virgin mouse. Taken by biopsy.;
MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,  
RA Alteschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska J., Smalish D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N;  
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old  
RC virgin mouse. Taken by biopsy;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC018179; AAH18179.1; -;  
DR HSP: Q93009; INBF.  
DR MEROPS; C19.019; -;  
DR MGD; MGI:2385198; Unpl.  
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.  
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001394; Peptidase\_C19.  
DR Pfam; PF00443; UCH; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
DR PROSITE; PS00972; UCH\_2\_1; 1.  
DR PROSITE; PS00973; UCH\_2\_2; 1.  
DR PROSITE; PS00235; UCH\_2\_3; 1.  
KW Protease.  
SQ SEQUENCE 784 AA; 87445 MW; F93699AF28F2B24F CRC64;  
Query Match 46.9%; Score 46; DB 2; Length 784;  
Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 RSDYKLYNNSTLKDILG 18  
:|||||:|:|:  
Db 673 KADYELYNKASNPDKVVG 690  
RESULT 34  
DPOL THEST STANDARD; PRT; 1829 AA.  
AC O33845;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE DNA polymerase (SC 2.7.7.7).  
GN Name=pol;  
OS Thermococcus sp. (strain TV).  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=110163;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98094267; PubMed=9434178;  
RA Niehaus F., Frey B., Antranikian G.;  
RT "Cloning and characterisation of a thermostable alpha-DNA polymerase  
RT from the hyperthermophilic archaeon Thermococcus sp. TY.";  
RL Gene 204:153-158 (1997).

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} (N).  
CC -1- PRT: This protein undergoes a protein self splicing that involves  
CC a post-translational excision of the three intervening regions  
CC (intense) followed by peptide ligation.  
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y13030; CAA73475.1; -;  
DR HSP; P77933; LGCX.  
DR InterPro; IPR006172; DNA\_pol\_B.  
DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
DR InterPro; IPR006134; DNA\_pol\_B\_region.  
DR InterPro; IPR006142; INTEIN.  
DR InterPro; IPR004042; Intein\_endonuc.  
DR InterPro; IPR006141; Intein\_S.  
DR InterPro; IPR004578; Pol2.  
DR Pfam; PF00136; DNA\_pol\_B; 4.  
DR Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
DR PRINTS; PR00379; INTEIN.  
DR TIGRFAMs; TIGR01443; Intein\_Cterm; 3.  
DR TIGRFAMs; TIGR01445; Intein\_Nterm; 3.  
DR TIGRFAMs; TIGR00592; Pol2; 2.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
DR PROSITE; PS00818; INTEIN\_C\_TER; 3.  
DR PROSITE; PS00819; INTEIN\_ENDONUCLEASE; 2.  
DR PROSITE; PS00817; INTEIN\_N\_TER; 3.  
KW Autocatalytic cleavage; DNA replication; DNA-binding;  
KW DNA-directed DNA polymerase; Endonuclease; Hydrolase;  
KW Protein splicing; Transferase.  
FT CHAIN 1 409 DNA polymerase, 1st part (Potential).  
FT CHAIN 410 769 Intein I.  
FT CHAIN 770 855 DNA polymerase, 2nd part (Potential).  
FT CHAIN 856 1392 Intein II.  
FT CHAIN 1393 1441 DNA polymerase, 3rd part (Potential).  
FT CHAIN 1442 1598 Intein III.  
FT CHAIN 1599 1829 DNA polymerase, 4th part (Potential).  
SQ SEQUENCE 1829 AA; 211875 MW; A113A8C57EB9CB3 CRC64;  
Query Match 46.9%; Score 46; DB 1; Length 1829;  
Best Local Similarity 61.5%; Pred. No. 5.9e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 5 KLYNKSSTLKDIL 17  
|||||:|:|:  
Db 1161 KLYNENPNVLKDM 1173  
RESULT 35  
Q7K5Q6 PRELIMINARY; PRT; 2006 AA.  
AC Q7K5Q6;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Erythrocyte binding protein 3.  
GN Name=maebi;  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22072559; PubMed=12076768;  
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;  
RT "Plasmodium falciparum MAEBL is a unique member of the ebl family.";  
RL Mol. Biochem. Parasitol. 122:35-44 (2002).

```

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22077637; PubMed=12082132;
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Evolutionary relationships of conserved cysteine-rich motifs in
RL adhesive molecules of malaria parasites.";
RN Mol. Biol. Evol. 19:1128-1142(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY042084; AAQ73469.1; -
SQ SEQUENCE 2006 AA; 237620 MW; EE75EP42E2BF767D CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2006;
Best Local Similarity 53.3%; Pred. No. 6.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 SDYKLYNKNSTLKD 16
Db 551 TDNKLNNNNQIKD 565

RESULT 36
AAQ73469
ID AAQ73469 PRELIMINARY; PRT; 2006 AA.
AC AAQ73469;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Erythrocyte binding protein 3.
GN MAEBL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22072559; PubMed=12076768;
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;
RT "Plasmodium falciparum MAEBL is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 122:35-44(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY042084; AAQ73469.1; -
SQ SEQUENCE 2006 AA; 237620 MW; EE75EP42E2BF767D CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2006;
Best Local Similarity 53.3%; Pred. No. 6.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 SDYKLYNKNSTLKD 16
Db 551 TDNKLNNNNQIKD 565

RESULT 37
AAQ73469
ID AAQ73469 PRELIMINARY; PRT; 2019 AA.
AC AAQ73469;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Erythrocyte binding protein 2.
GN MAEBL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22072559; PubMed=12076768;
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;
RT "Evolutionary relationships of conserved cysteine-rich motifs in
RL adhesive molecules of malaria parasites.";
RN Mol. Biol. Evol. 19:1128-1142(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY042084; AAQ73469.1; -
SQ SEQUENCE 2006 AA; 237620 MW; EE75EP42E2BF767D CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2006;
Best Local Similarity 53.3%; Pred. No. 6.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 SDYKLYNKNSTLKD 16
Db 551 TDNKLNNNNQIKD 565

RESULT 38
AAQ73468
ID AAQ73468 PRELIMINARY; PRT; 2019 AA.
AC AAQ73468;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Erythrocyte binding protein 2.
GN MAEBL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22072559; PubMed=12076768;
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;
RT "Plasmodium falciparum MAEBL is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 122:35-44(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY042084; AAQ73468.1; -
SQ SEQUENCE 2019 AA; 239410 MW; A626F2C684C08785 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2019;
Best Local Similarity 53.3%; Pred. No. 6.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 SDYKLYNKNSTLKD 16
Db 551 TDNKLNNNNQIKD 565

RESULT 39
AAQ73468
ID AAQ73468 PRELIMINARY; PRT; 2019 AA.
AC AAQ73468;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Erythrocyte binding protein 2.
GN MAEBL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22072559; PubMed=12076768;
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;
RT "Evolutionary relationships of conserved cysteine-rich motifs in
RL adhesive molecules of malaria parasites.";
RN Mol. Biol. Evol. 19:1128-1142(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY042084; AAQ73468.1; -
SQ SEQUENCE 2019 AA; 239410 MW; A626F2C684C08785 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2019;
Best Local Similarity 53.3%; Pred. No. 6.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 SDYKLYNKNSTLKD 16
Db 551 TDNKLNNNNQIKD 565

```

```
RESULT 39
Q81HP3 ID Q81HP3 PRELIMINARY; PRT; 2055 AA.
AC Q81HP3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MAEBL, putative.
GN ORFNames=PF11_0486;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pextea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014843; AAN36066.1; -.
DR HSSP; P04268; 1IC2.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008602; Duffy_binding.
DR Pfam; PF05424; Duffy_binding; 1.
SQ SEQUENCE 2055 AA; 243211 MW; 783F48BB7286B1F3 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2055;
Best Local Similarity 53.3%; Pred. No. 6.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNNKNSSTLKD 16
Db 551 TDNKLNNNNQIKD 565

RESULT 40
Q8T5C7 ID Q8T5C7 PRELIMINARY; PRT; 2055 AA.
AC Q8T5C7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Erythrocyte binding protein 1 (Chimeric erythrocyte-binding protein
MAEBL).
GN Name=maeb1; Synonyms=MAEBL;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22072559; PubMed=12076768;
RA Blair P.L., Kappe S.H., Maciel J.E., Balu B., Adams J.H.;
RT "Plasmodium falciparum MAEBL is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 122:35-44(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22077637; PubMed=12082132;
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Evolutionary relationships of conserved cysteine-rich motifs in
adhesive molecules of malaria parasites.";
RL Mol. Biol. Evol. 19:1128-1142(2002).
RN [3]
```

```
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22155635; PubMed=12165387;
RA Ghai M., Dutta S., Hall T., Freilich D., Ockenhouse C.;
RT "Identification, expression, and functional characterization of MAEBL,
a sporozoite and asexual blood stage chimeric erythrocyte-binding
protein of Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 123:35-45(2002).
DR EMBL; AY042084; AAL10509.1; -.
DR EMBL; AF400002; AAM90625.1; -.
DR HSSP; P04268; 1IC2.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008602; Duffy_binding.
DR Pfam; PF05424; Duffy_binding; 1.
SQ SEQUENCE 2055 AA; 243211 MW; 783F48BB7286B1F3 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 2055;
Best Local Similarity 53.3%; Pred. No. 6.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNNKNSSTLKD 16
Db 551 TDNKLNNNNQIKD 565

Search completed: November 24, 2004, 09:28:58
Job time : 90.3587 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:44 ; Search time 15.6957 Seconds  
(without alignments)  
116.473 Million cell updates/sec

Title: US-09-719-379A-4  
Perfect score: 98  
Sequence: 1 RSDYKLYNKNSSTLKDLGE 19  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: Pirl1.\*  
2: Pirl2.\*  
3: Pirl3.\*  
4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 87    | 88.8        | 353    | 2 C64187 | outer membrane pro |
| 2          | 49    | 50.0        | 167    | 2 A60013 | orf1 protein - fow |
| 3          | 49    | 50.0        | 909    | 1 A54809 | disease resistance |
| 4          | 48    | 49.0        | 430    | 2 T28318 | ORF MSV157 hypothe |
| 5          | 48    | 49.0        | 505    | 1 S77034 | protein kinase pkn |
| 6          | 46    | 46.9        | 465    | 2 A81419 | hypothetical prote |
| 7          | 45    | 45.9        | 1017   | 2 G85055 | probable polypote  |
| 8          | 44.5  | 45.4        | 315    | 2 B90594 | holliday junction  |
| 9          | 44    | 44.9        | 262    | 2 T49066 | hypothetical prote |
| 10         | 44    | 44.9        | 1058   | 2 S50295 | hypothetical prote |
| 11         | 44    | 44.9        | 1216   | 2 S46177 | probable Ca2+-tran |
| 12         | 43    | 43.9        | 131    | 2 A69824 | hypothetical prote |
| 13         | 43    | 43.9        | 148    | 2 C84164 | hypothetical prote |
| 14         | 43    | 43.9        | 241    | 2 S52703 | hypothetical prote |
| 15         | 43    | 43.9        | 252    | 2 H70600 | uncharacterized co |
| 16         | 43    | 43.9        | 263    | 2 H71854 | hypothetical prote |
| 17         | 43    | 43.9        | 342    | 2 B70100 | beta-N-acetylhexos |
| 18         | 43    | 43.9        | 429    | 2 S59773 | 26S proteasome reg |
| 19         | 43    | 43.9        | 609    | 2 H97040 | ATP-dependent DNA  |
| 20         | 43    | 43.9        | 625    | 2 H70330 | hypothetical prote |
| 21         | 42.5  | 43.4        | 645    | 1 WNV74  | p74 protein - Auto |
| 22         | 42.5  | 43.4        | 645    | 2 C72867 | occlusion-derived  |
| 23         | 42    | 42.9        | 176    | 2 JH0607 | crystsin inhibitor |
| 24         | 42    | 42.9        | 262    | 2 D90400 | hypothetical prote |
| 25         | 42    | 42.9        | 274    | 2 B70020 | conserved hypotet  |
| 26         | 42    | 42.9        | 292    | 2 F84951 | hypothetical prote |
| 27         | 42    | 42.9        | 319    | 1 OWBS   | ornithine carbamoy |
| 28         | 42    | 42.9        | 431    | 2 T38471 | probable chromatin |
| 29         | 42    | 42.9        | 589    | 2 S60159 | serine/threonine-s |

|    |      |      |      |          |                    |
|----|------|------|------|----------|--------------------|
| 30 | 42   | 42.9 | 591  | 2 T48596 | ankyrin-like prote |
| 31 | 42   | 42.9 | 739  | 2 E97245 | beta-lactamase cla |
| 32 | 42   | 42.9 | 785  | 2 T17309 | probable ubiquitin |
| 33 | 42   | 42.9 | 831  | 2 T25827 | hypothetical prote |
| 34 | 42   | 42.9 | 947  | 2 T08605 | hypothetical prote |
| 35 | 42   | 42.9 | 1035 | 2 AD3203 | autotransporter pr |
| 36 | 42   | 42.9 | 1325 | 2 T42722 | male-enhanced anti |
| 37 | 42   | 42.9 | 1498 | 2 E97355 | DNA segregation AT |
| 38 | 42   | 42.9 | 1553 | 2 S67483 | adenosinetriphosph |
| 39 | 41.5 | 42.3 | 615  | 2 E86212 | hypothetical prote |
| 40 | 41.5 | 42.3 | 1464 | 2 S47555 | N-methyl-D-asparta |
| 41 | 41   | 41.8 | 125  | 2 F70106 | hypothetical prote |
| 42 | 41   | 41.8 | 152  | 2 D86648 | ferrous ion transp |
| 43 | 41   | 41.8 | 193  | 2 E71713 | hypothetical prote |
| 44 | 41   | 41.8 | 270  | 2 E69996 | amino acid ABC tra |
| 45 | 41   | 41.8 | 351  | 2 F72020 | leucine dehydrogen |

ALIGNMENTS

RESULT 1

C64187  
outer membrane protein A homolog - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: C64187  
R:Flieischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.  
R:Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
R:D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: C64187  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-353 <TIGR>  
A:Cross-references: UNIPROT:P43840; GB:U32796; GB:L42023; NID:g1574086; PIDN:AAC22819.1;  
C:Superfamily: outer membrane protein A

Query Match 88.8%; Score 87; DB 2; Length 353;  
Best Local Similarity 89.5%; Pred. No. 5.4e-06;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSTLKDLGE 19  
|||||:|||||  
Db 132 RSDYKLYNKNSSTLKDLGE 150

RESULT 2

A60013  
orf1 protein - fowlpox virus  
C:Species: fowlpox virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C:Accession: A60013  
R:Kumar, S.; Boyle, D.B.  
Virus Res. 15, 175-186, 1990  
A:Title: Mapping of a major early/late gene of fowlpox virus.  
A:Reference number: A60013; MUID:90209363; PMID:2157318  
A:Accession: A60013  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-167 <KUM>  
A:Cross-references: UNIPROT:Q05569; GB:X52461; NID:g62320; PIDN:CAA36694.1; PID:g62321  
C:Superfamily: RNA polymerase subunit 19, poxvirus type

Query Match 50.0%; Score 49; DB 2; Length 167;  
Best Local Similarity 38.9%; Pred. No. 2.5;  
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSTLKDLGE 19  
:|||||:::|

Db 16 NDYKSYDEDDSDSGICE 33

RESULT 3

A54809

disease resistance protein RPS2 - Arabidopsis thaliana

N;Alternate names: protein F20B18.200

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C;Accession: A54809; A54811; T04264

R;Mindrinosa, M.; Katagiri, F.; Yu, G.L.; Ausubel, F.M.

Cell 78, 1089-1099, 1994

A;Title: The Arabidopsis thaliana disease resistance gene RPS2 encodes a protein containing

A;Reference number: A54809; MUID:95007758; PMID:7923358

A;Accession: A54809

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-909 <MIN>

A;Cross-references: UNIPROT:Q42484; GB:U12860; NID:G549978; PIDN:AAAS0236.1; PID:G549978

R;Bent, A.F.; Kunkel, B.N.; Dahlbeck, D.; Brown, K.L.; Schmidt, R.; Giraudat, J.; Leung, S.

Science 265, 1856-1860, 1994

A;Title: RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance

A;Reference number: A54811; MUID:94377978; PMID:8091210

A;Accession: A54811

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA; mRNA

A;Residues: 1-909 <BEN>

A;Cross-references: GB:U14158; NID:G548085; PIDN:AAA21874.1; PID:G548086

R;Bavan, M.; Rose, M.; Hempel, S.; Etanian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 1999

A;Reference number: 215263

A;Accession: T04264

A;Molecule type: DNA

A;Residues: 1-909 <BEV>

A;Cross-references: EMBL:AL049483

A;Experimental source: cultivar Columbia; BAC clone F20B18

C;Genetics:

A;Gene: RPS2

A;Map position: 4

A;Note: F20B18.200

C;Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

C;Keywords: leucine zipper; membrane protein; nucleotide binding

Query Match 50.0%; Score 49; DB 1; Length 909;  
Best Local Similarity 55.6%; Pred. No. 15;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDLGE 19  
:|||||:|:|:|:|:  
Db 111 ADYKLCCKVSAILKSGICE 128

RESULT 4

T28318

ORF MSV157 hypothetical protein - Melanoplus sanguinipes entomopoxvirus

C;Species: Melanoplus sanguinipes entomopoxvirus

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004

C;Accession: T28318

R;Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.

J. Virol. 73, 533-552, 1999

A;Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A;Reference number: Z20484; MUID:99102612; PMID:9847359

A;Accession: T28318

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-430 <AFO>

A;Cross-references: UNIPROT:Q9VVT5; EMBL:AF063866; NID:G4049647; PIDN:AAAC97678.1; PID:G4049647

C;Genetics:

A;Note: MSV157

Query Match 49.0%; Score 48; DB 2; Length 430;  
Best Local Similarity 69.2%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

probable polyprotein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: G85055  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617198  
A:Accession: G85055  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1017 <STO>  
A:Cross-references: UNIPROT:Q9XEB9; GB:NC\_001268; NID:g7267198; PIDN:CAB77909.1; GSPDB:C  
C:Genetics:  
A:Map position: 4

Query Match 45.9%; Score 45; DB 2; Length 1017;  
Best Local Similarity 52.9%; Pred. No. 74;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKDLGE 19  
|||:::|||||  
Db 936 DTKLFLKSTFDIKDLGE 952

RESULT 8  
B90594  
holliday junction DNA helicase ruvb [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C:Accession: B90594  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: B90594  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-315 <KUR>  
A:Cross-references: UNIPROT:Q98PR1; GB:AL445566; PID:g14090073; PIDN:CAC13831.1; GSPDB:C  
C:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPU 6580  
A:Genetic code: SGC3  
C:Superfamily: ruvB protein

Query Match 45.4%; Score 44.5; DB 2; Length 315;  
Best Local Similarity 68.8%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

Qy 3 DYKLYNKSSTLKDLG 18  
|||||  
Db 227 DYKLVNK---TLXQLG 239

RESULT 9  
T49066  
hypothetical protein T5P19.300 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: T49066  
R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansoorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.;  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25014  
A:Accession: T49066  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <BEN>  
A:Cross-references: UNIPROT:Q9LXX5; EMBL:AL163972; GSPDB:GN00061; ATSP:T5P19.300  
C:Genetics:  
A:Experimental source: cultivar Columbia; BAC clone T5P19  
A:Gene: ATSP:T5P19.300

A:Map position: 3  
A:Introns: 101/1; 206/3

Query Match 44.9%; Score 44; DB 2; Length 262;  
Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KLYNKSSTLKDLGE 19  
:|||:::|||||  
Db 158 RLTKNPNATLDELGE 172

RESULT 10  
S50295  
hypothetical protein YKR028w homolog YJL098w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J0840  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Aug-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S50295; S53389; S56875; S56876; S57370  
R:Mosga, T.; Boles, E.; Schaeff-Gerstenschlaeger, I.; Schmitt, S.; Zimmermann, P.K.  
Yeast 10, 1481-1488, 1994  
A:Title: Sequence and function analysis of a 9.74 kb fragment of Saccharomyces cerevisia  
A:Reference number: S50295; MUID:95176706; PMID:7871887  
A:Accession: S50295  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1058 <MIO>  
A:Cross-references: UNIPROT:P40856; EMBL:X77923; NID:g640004; PIDN:CAAS4892.1; PID:g6400  
R:Rasmussen, S.W.  
submitted to the EMBL Data Library, February 1995  
A:Description: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and  
A:Reference number: S53376  
A:Accession: S53389  
A:Molecule type: DNA  
A:Residues: 1-852 <RAS>  
A:Cross-references: EMBL:X85021  
R:Mosga, T.; Schaeff-Gerstenschlaeger, I.; Baur, A.; Boles, E.; Chaltatzis, N.; Fournie  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56875  
A:Accession: S56875  
A:Molecule type: DNA  
A:Residues: 1-1058 <MIF>  
A:Cross-references: EMBL:Z49373; NID:g1008275; PIDN:CAA89392.1; PID:g1008276; MIPS:YJL09  
R:Rasmussen, S.W.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56876  
A:Accession: S56876  
A:Molecule type: DNA  
A:Residues: 1-852 <RAW>  
A:Cross-references: EMBL:Z49373; MIPS:YJL098w  
R:Rasmussen, S.W.  
Yeast 11, 873-883, 1995  
A:Title: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3 g  
A:Reference number: S57357; MUID:96090136; PMID:7483851  
A:Accession: S57370  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-852 <RAF>  
A:Cross-references: EMBL:X85021  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995  
C:Genetics:  
A:Gene: SGD:SAP185  
A:Cross-references: SGD:S0003634  
A:Map position: 10L

Query Match 44.9%; Score 44; DB 2; Length 1058;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKDL 17  
|||||  
Db 72 DYKAYKPNLSLNDL 86



R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-596, 1997

A:Authors: Smith, H.O.; Ventner, J.C.

A>Title: Genomic sequence of a Lyme disease spirochaste, Borrelia burgdorferi.

A:Reference number: A70100; PMID:98065943; PMID:9403685

A:Accession: B70100

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-342 <KLE>

A:Cross-references: UNIPROT:O54536; GB:AE001115; GB:AE000783; NID:g2687879; PIDN:AAC6640

A:Experimental source: strain B31

Query Match 43.9%; Score 43; DB 2; Length 342;  
Best Local Similarity 53.3%; Pred. No. 49;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKDL 17  
||: ||: | |||  
Db 29 DYEYFNKDSDLVDL 43

RESULT 18

S59773

26S proteasome regulatory particle chain RPN7 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein P8283.8; protein YPR108w

C:Species: Saccharomyces cerevisiae

C>Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004

C:Accession: S59773

R:Nelson, J.

A:Submitted to the EMBL Data Library, July 1995

A:Description: The sequence of S. cerevisiae cosmid 8283.

A:Reference number: S59764

A:Accession: S59773

A:Molecule type: DNA

A:Residues: 1-429 <NEL>

A:Cross-references: UNIPROT:Q06103; EMBL:U32445; NID:g914969; PIDN:AAB68078.1; PID:g9149

C:Genetics:

A:Gene: SGD:RPN7; RPN7; MIPS:YPR108w

A:Cross-references: MIPS:YPR108w; SGD:S0006312

A:Map position: 16R

C:Keywords: nucleotide binding; P-loop

F:150-157/Region: nucleotide-binding motif A (P-loop)

Query Match 43.9%; Score 43; DB 2; Length 429;  
Best Local Similarity 47.6%; Pred. No. 62;  
Matches 10; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

Qy 3 DYKLYN----KNSSSTLKDLGE 19  
| : ||| | || | : | : |  
Db 95 DQELYNELCKNESKIKELNE 115

RESULT 19

H97040

ATP-dependent DNA helicase, uvrB-like [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004

C:Accession: H97040

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC8224

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: H97040

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-609 <KUR>

A:Cross-references: UNIPROT:Q97JY0; GB:AE001437; PIDN:AAK79115.1; PID:g15024062; GSPDB:B:GSPDB:B:1

A:Experimental source: Clostridium acetobutylicum ATCC8224

C:Genetics:

A:Gene: CAC1142



C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 16-Aug-2004  
C;Accession: D90400  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, E.  
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: D90400  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-262 <KUR>  
A;Cross-references: UNIPROT:Q97WC9; GB:AEO06641; MID:gl3815601; PIDN:AAK42459.1; GSPDB:B  
C;Genetics:  
A;Gene: cbif  
C;Superfamily: Precorrin-4 C11-methyltransferase

Query Match 42.9%; Score 42; DB 2; Length 262;  
Best Local Similarity 42.1%; Pred.No. 53;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 RSDVKLYKNKSSTLKOLGE 19  
          |::|||:|||::|  
Db 48 RKOREVYNSSSLTNEIVE 66  
          |::|||:|||::|

RESULT 25  
B70020  
conserved hypothetical protein yueA - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: B70020  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet-  
C.; Bron, S.; Brouillet, S.; Bruschii, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd-  
A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier-  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.P.  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mauel-  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle-  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron-  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpesto, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A;Authors: Yoshihawa, H.F.; Zumstein, E.; Yoshihawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: B70020  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A;Residues: 1-274 <KUN>  
A;Cross-references: UNIPROT:O32167; GB:Z99120; GB:AL009126; MID:g2635613; PIDN:CAB15262.  
C;Genetics:  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: yusa  
C;Superfamily: lipoprotein-28

Query Match 42.9%; Score 42; DB 2; Length 274;  
Best Local Similarity 87.5%; Pred.No. 55;  
Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 SDVKLYNK 9  
          |::|||:|||  
Db 66 SDYKMYNK 73  
          |::|||:|||

RESULT 26  
F84951  
hypothetical protein yfiB [imported] - Buchnera sp. (strain APS)  
C;Species: Buchnera sp.  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text change 16-Aug-2004

C;Genetics:  
A;Gene: argP  
A;Map position: 100 (degrees)  
C;Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase  
C;Keywords: arginine biosynthesis; transferase  
F1-319/Product: ornithine carbamoyltransferase #status experimental <MAT>  
F12-310/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 42.9%; Score 42; DB 1; Length 319;  
Best Local Similarity 64.3%; Pred. No. 65;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 6 LYNNKNSSTLKDLGE 19  
|||: |||||  
Db 9 LYGKLLTLKDLSE 22

RESULT 28  
TJ38471  
probable chromatin assembly factor - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Aug-2004  
C;Accession: TJ38471  
R;Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1997  
A;Reference number: T21796  
A;Accession: TJ38471  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-431 <BRO>  
A;Cross-references: UNIPROT:O14021; EMBL:Z97210; PIDN:CAB10144.1; GSPDB:GN000066; SPDB:SE  
A;Experimental source: strain 972h-; cosmid c29A4  
C;Genetics:  
A;Gene: SPDB:SPAC29A4.18  
A;Map position: 1  
A;Introns: 29/1; 64/2  
C;Superfamily: WD repeat homology

Query Match 42.9%; Score 42; DB 2; Length 431;  
Best Local Similarity 53.3%; Pred. No. 89;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKDL 17  
:||: |||||  
Db 29 EYKIWKNSPFLYDL 43

RESULT 29  
S60159  
serine/threonine-specific kinase (EC 2.7.1.-) isoform HST7-T - yeast (Candida albicans)  
C;Species: Candida albicans  
C;Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 16-Aug-2004  
C;Accession: S60159  
R;Clark, K.L.; Feldmann, P.J.F.; Dignard, D.; Larocque, R.; Brown, A.J.P.; Lee, M.G.; Th  
Mol. Gen. Genet. 249, 609-621, 1995  
A;Title: Constitutive activation of the Saccharomyces cerevisiae mating response pathway  
A;Reference number: S60154; MUID:96132546; PMID:8544826  
A;Accession: S60159  
A;Molecule type: DNA  
A;Residues: 1-589 <CLA>  
A;Cross-references: EMBL:L19195  
A;Experimental source: isoform HST7-T  
C;Superfamily: protein kinase homology  
C;Keywords: ATP; phosphotransferase; protein kinase  
F;247-564/Domain: protein kinase homology <KIN>  
F;255-263/Region: protein kinase ATP-binding motif

Query Match 42.9%; Score 42; DB 2; Length 589;  
Best Local Similarity 58.8%; Pred. No. 1.2e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDLG 18  
| |||||

Db 240 SSLKLNKDLTLTKQLG 256

RESULT 30  
T48596  
ankyrin-like protein - Arabidopsis thaliana  
N;Alternate names: protein F18022.20  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C;Accession: T48596  
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lee  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24493  
A;Accession: T48596  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-591 <BEV>  
A;Cross-references: UNIPROT:Q9LYA6; EMBL:DB  
A;Experimental source: cultivar Columbia; BAC clone F18022  
C;Genetics:  
A;Map position: 5  
A;Introns: 32/3; 78/3  
A;Note: F18022.20

Query Match 42.9%; Score 42; DB 2; Length 591;  
Best Local Similarity 44.4%; Pred. No. 1.2e+02;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSTLKDLGE 19  
:|||||: |||  
Db 342 ADVKLYNNSGDTVVSLSE 359

RESULT 31  
E97245  
beta-lactamase class C domain (BBPX family) containing protein [imported] - Clostridium  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: E97245  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: E97245  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-739 <KUR>  
A;Cross-references: UNIPROT:Q97FD2; GB:AE001437; PIDN:AAK80752.1; PID:gl5025849; GSPDB:G  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2808

Query Match 42.9%; Score 42; DB 2; Length 739;  
Best Local Similarity 63.6%; Pred. No. 1.6e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSNS 11  
:|||||: |||  
Db 621 QDDYKLYKKNN 631

RESULT 32  
T17309  
probable ubiquitin thiolesterase (EC 3.1.2.15) UBP [similarity] - human  
N;Alternate names: protein DKFZp434A028.1; protein DKFZp434E0635.1  
C;Species: Homo sapiens (man)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T17309; T17277  
R;Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A;Reference number: Z18726  
A;Accession: T17309



```

Query Match      42.9%; Score 42; DB 2; Length 947;
Best Local Similarity 53.8%; Pred. No. 2.1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      7 YNKNSSTLKDLGE 19
Db      287 YNKNNNIDDLSE 299
      ||||:..|||
      ||||:..|||

RESULT 35
AD3203
A:Title: autotransporter protein bapA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD3203
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3203
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1035 <R>
A:Cross-references: UNIPROT:Q8UJX1; GB:AE008687; PIDN:AAL46042.1; PID:g17743801; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: bapA
A:Genome: plasmid

Query Match      42.9%; Score 42; DB 2; Length 1035;
Best Local Similarity 46.7%; Pred. No. 2.3e+02;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy      3 DYKLYNKNSSTLKDL 17
Db      267 DYGYSRNASSATDL 281
      |||:|:|:|
      |||:|:|:|

RESULT 36
T42722
A:Title: male-enhanced antigen-2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42722
R:Kondo, M.; Sutou, S.
DNA Seq. 7, 71-82, 1997
A:Title: Cloning and molecular characterization of cDNA encoding a mouse male-enhanced a
A:Reference number: Z2242; MUID:97217683; PMID:9063644
A:Accession: T42722
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1325 <R>
A:Cross-references: UNIPROT:P59337; EMBL:D78270; NID:d1096175; PID:d1020389; PIDN:BAAL196
A:Experimental source: strain CD-1
C:Function:
A:Description: supposed to play some role for spermatogenesis
C:Keywords: leucine zipper

Query Match      42.9%; Score 42; DB 2; Length 1325;
Best Local Similarity 47.4%; Pred. No. 2.9e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy      1 RSDYKLYNKNSSTLKDLGE 19
Db      1307 RGDTKLHNQNSVPRDLGQ 1325
      ||||:|:|:|
      ||||:|:|:|

RESULT 37

```

DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Close  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: B97355  
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97355  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1498 <KUR>  
A:Cross-references: UNIPROT:Q04351; GB:AE001437; PIDN:AAK81629.1; PID:g15026814; GSPDB:G15026814  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3709

Query Match 42.9%; Score 42; DB 2; Length 1498;  
Best Local Similarity 44.4%; Pred. No. 3.3e+02;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSTLKDGE 19  
|:::| | | | |  
Db 487 SFYQILNKRKQVLKDAGK 504

RESULT 38  
S67483  
adenosinetriphosphatase 2 - malaria parasite (Plasmodium falciparum)  
N:Alternate names: ATPase 2  
C:Species: Plasmodium falciparum  
C:Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 09-Jul-2004  
C:Accession: S67483  
R:Trottien, F.; Cowman, A.F.  
Eur. J. Biochem. 227, 214-225, 1995  
A:Title: Molecular cloning and sequence of two novel P-type adenosinetriphosphatases from Plasmodium falciparum  
A:Reference number: S67483; MUID:95154293; PMID:7851389  
A:Accession: S67483  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1553 <TRO>  
A:Cross-references: UNIPROT:Q9U421; UNIPROT:Q815L4; EMBL:U16955  
C:Genetics:  
A:Introns: 17/2

Query Match 42.9%; Score 42; DB 2; Length 1553;  
Best Local Similarity 42.1%; Pred. No. 3.5e+02;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSTLKDGE 19  
|:::| | | | |  
Db 349 KSDIKYNNKDDDFDDMD 367

RESULT 39  
E86212  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: E86212  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Hughes, B.; Huizar, L.; Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86212

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-615 <STO>  
A:Cross-references: UNIPROT:Q9LQP9; GB:AE005172; NID:g84339895; PIDN:AAF75081.1; GSPDB:G15026814  
C:Genetics:  
A:Map position: 1  
C:Superfamily: tomato leucine zipper-containing protein

Query Match 42.3%; Score 41.5; DB 2; Length 615;  
Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
Matches 10; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Qy 2 SDYKLYNKNSTLKDGE 18  
|:::| | | | |  
Db 152 SDGELYSESSSELGKDG 169

RESULT 40  
S47555  
N-methyl-D-aspartate receptor modulatory chain hNR2A - human  
C:Species: Homo sapiens (man)  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C:Accession: S47555  
R:Foldes, R.L.; Adams, S.L.; Fantiske, R.P.; Kamboj, R.K.  
Biochim. Biophys. Acta 1223, 155-159, 1994  
A:Title: Human N-methyl-D-aspartate receptor modulatory subunit hNR2A: cloning and sequencing of the complementary DNA  
A:Reference number: S47555; MUID:94339179; PMID:8061049  
A:Accession: S47555  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1464 <FOL>  
A:Cross-references: UNIPROT:Q12879; GB:U09002; NID:g558748; PIDN:AAE60343.1; PID:g558749  
C:Superfamily: N-methyl-D-aspartate receptor 2A; glutamate receptor homology F; 428-854/Domain: glutamate receptor homology <GRH>

Query Match 42.3%; Score 41.5; DB 2; Length 1464;  
Best Local Similarity 66.7%; Pred. No. 3.9e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 4 YKLYNKNSTLKDGE 18  
|:::| | | | |  
Db 1184 YKLYSKH-FTLKDKG 1197

Search completed: November 24, 2004, 09:30:19  
Job time : 17.6957 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:04 ; Search time 101.152 Seconds  
(without alignments)  
125.141 Million cell updates/sec

Title: US-09-719-379A-2  
Perfect score: 112  
Sequence: 1 RSDYKLYNKSSNSTLKNLGE 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description         |
|------------|-------|---------------|--------|--------------|---------------------|
| 1          | 76.5  | 68.3          | 353    | 1 OM51_HAEIN | P43840 haemophilus  |
| 2          | 56    | 50.0          | 168    | 2 Q939P1     | Q939P1 lactococcus  |
| 3          | 53    | 47.3          | 442    | 1 TIG_BUCAI  | P57546 buchnera ap  |
| 4          | 52.5  | 46.9          | 350    | 2 Q8STT1     | Q8STT1 cephalito    |
| 5          | 52    | 46.4          | 322    | 2 Q82CA4     | Q82CA4 streptomyce  |
| 6          | 51    | 45.5          | 322    | 2 Q832V3     | Q832V3 enterococcu  |
| 7          | 50.5  | 45.1          | 907    | 2 Q8LL02     | Q8LL02 arabidopsis  |
| 8          | 50    | 44.6          | 287    | 2 Q6L0G9     | Q6L0G9 microphilus  |
| 9          | 50    | 44.6          | 369    | 2 Q7RA98     | Q7RA98 plasmodium   |
| 10         | 50    | 44.6          | 418    | 2 Q7RL88     | Q7RL88 plasmodium   |
| 11         | 49.5  | 44.2          | 244    | 2 Q96Z64     | Q96Z64 sulfolobus   |
| 12         | 49    | 43.8          | 351    | 2 Q6GUB7     | Q6GUB7 pasteurilla  |
| 13         | 49    | 43.8          | 484    | 2 Q9ALF8     | Q9ALF8 campylobact  |
| 14         | 49    | 43.8          | 502    | 2 Q7RF20     | Q7RF20 plasmodium   |
| 15         | 48.5  | 43.3          | 150    | 2 Q22762     | Q22762 caenorhabdi  |
| 16         | 48.5  | 43.3          | 5561   | 2 Q8ILC9     | Q8ILC9 plasmodium   |
| 17         | 48    | 42.9          | 245    | 2 Q7RHN4     | Q7RHN4 plasmodium   |
| 18         | 48    | 42.9          | 384    | 2 Q9CUF6     | Q9CUF6 mus musculu  |
| 19         | 48    | 42.9          | 409    | 2 Q74BR0     | Q74BR0 geobacter s  |
| 20         | 48    | 42.9          | 409    | 2 AAR35357   | AAR35357 geobacter  |
| 21         | 48    | 42.9          | 706    | 2 Q8ZJD9     | Q8ZJD9 versinia pe  |
| 22         | 48    | 42.9          | 706    | 2 AAS60450   | AAS60450 versinia   |
| 23         | 48    | 42.9          | 952    | 2 Q8MRNS     | Q8MRNS drosophila   |
| 24         | 48    | 42.9          | 1108   | 2 Q8T148     | Q8T148 dictyosteli  |
| 25         | 48    | 42.9          | 6761   | 2 Q6LEX8     | Q6LEX8 plasmodium   |
| 26         | 48    | 42.9          | 6761   | 2 CAG25109   | CAG25109 plasmodium |
| 27         | 47.5  | 42.4          | 438    | 1 ARLY_CLOTE | P59616 clostridium  |
| 28         | 47.5  | 42.4          | 595    | 2 Q7RLW6     | Q7RLW6 plasmodium   |
| 29         | 47    | 42.0          | 200    | 1 HRL2_LACNU | P22796 lachesis mu  |
| 30         | 47    | 42.0          | 229    | 2 Q7K6A7     | Q7K6A7 plasmodium   |
| 31         | 47    | 42.0          | 229    | 2 Q8T357     | Q8T357 plasmodium   |

|    |      |      |      |   |            |                    |
|----|------|------|------|---|------------|--------------------|
| 32 | 47   | 42.0 | 246  | 2 | Q6DKA7     | O6dka7 xenopus lae |
| 33 | 47   | 42.0 | 308  | 2 | Q50827     | Q50827 borrelia bu |
| 34 | 47   | 42.0 | 426  | 2 | Q6L273     | Q6l273 methanococc |
| 35 | 47   | 42.0 | 426  | 2 | CAF30312   | Caf30312 methanoco |
| 36 | 47   | 42.0 | 510  | 2 | Q8W2N4     | Q8w2n4 nicotiana t |
| 37 | 47   | 42.0 | 842  | 2 | Q8DTD8     | Q8dtd8 streptococc |
| 38 | 47   | 42.0 | 909  | 2 | Q756H9     | Q756h9 ashbya gose |
| 39 | 47   | 42.0 | 909  | 2 | AAS52963   | Aas52963 ashbya go |
| 40 | 47   | 42.0 | 1078 | 2 | Q7RCN2     | Q7rcn2 plasmodium  |
| 41 | 47   | 42.0 | 1374 | 1 | YMN3_YEAST | Q03099 saccharomyc |
| 42 | 47   | 42.0 | 1612 | 2 | Q6BTM5     | Q6btm5 debaryomyce |
| 43 | 47   | 42.0 | 2031 | 2 | Q7RF86     | Q7rf86 plasmodium  |
| 44 | 46.5 | 41.5 | 1262 | 2 | Q868L5     | Q868l5 plasmodium  |
| 45 | 46.5 | 41.5 | 1262 | 2 | Q8ID44     | Q8id44 plasmodium  |

#### ALIGNMENTS

##### RESULT 1

| ID | OM51_HAEIN   | STANDARD; | PRT; | 353 AA. |
|----|--|-----------|------|---------|
| AC | P43840;  |           |      |         |
| DT | 01-NOV-1995 (Rel. 32, Created)   |           |      |         |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update)  |           |      |         |
| DT | 01-OCT-2004 (Rel. 45, Last annotation update)  |           |      |         |
| DB | Outer membrane protein P5 precursor (OMP P5).  |           |      |         |
| GN | Names=ompA; Synonyms=ompP5; OrderedLocusNames=H11164;  |           |      |         |
| OS | Haemophilus influenzae.  |           |      |         |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;   |           |      |         |
| OC | Pasteurellaceae; Haemophilus.  |           |      |         |
| OX | NCBI_TaxID=727;  |           |      |         |
| RN | [1]  |           |      |         |
| RP | SEQUENCE FROM N.A.   |           |      |         |
| RC | STRAIN=Rd / KW20 / ATCC 51907;   |           |      |         |
| RX | MEDLINE=95350630; PubMed=7542800;  |           |      |         |
| RA | Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,   |           |      |         |
| RA | McKenney K., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  |           |      |         |
| RA | Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,   |           |      |         |
| RA | Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  |           |      |         |
| RA | Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,   |           |      |         |
| RA | Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  |           |      |         |
| RA | Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  |           |      |         |
| RA | Venter J.C.;   |           |      |         |
| RT | "Whole-genome random sequencing and assembly of Haemophilus influenzae   |           |      |         |
| RT | Rd."   |           |      |         |
| RL | Science 269:496-512 (1995).  |           |      |         |
| CC | -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.   |           |      |         |
| CC | -1- SIMILARITY: Belongs to the ompA family.  |           |      |         |
| CC | -----  |           |      |         |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration   |           |      |         |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |           |      |         |
| CC | the European Bioinformatics Institute. There are no restrictions on its  |           |      |         |
| CC | use by non-profit institutions as long as its content is in no way   |           |      |         |
| CC | modified and this statement is not removed. Usage by and for commercial  |           |      |         |
| CC | entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |           |      |         |
| CC | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |           |      |         |
| CC | -----  |           |      |         |
| CC | EMBL; U32796; AAC22819.1; --   |           |      |         |
| DR | PIR; C64187; C64187.   |           |      |         |
| DR | HSSP; P02934; 1QJP.  |           |      |         |
| DR | TIGR; H11164; --   |           |      |         |
| DR | InterPro; IPR006664; Bac OmpA.   |           |      |         |
| DR | InterPro; IPR002368; OmpA.   |           |      |         |
| DR | InterPro; IPR006665; OmpA/MotB.  |           |      |         |
| DR | InterPro; IPR006690; OmpA_LINK.  |           |      |         |
| DR | InterPro; IPR000498; OmpA_tmam.  |           |      |         |
| DR | Pfam; PF00691; OmpA; 1.  |           |      |         |
| DR | Pfam; PF01389; OmpA membrane; 1.   |           |      |         |
| DR | PRINTS; PR01021; OMPADOMAIN.   |           |      |         |
| DR | ProDom; PD000930; OmpA/MotB; 1.  |           |      |         |
| DR | PROSITE; PS01068; OmpA; 1.   |           |      |         |

```

KW Complete proteome; Outer membrane; Porin; Signal; Transmembrane.
FT SIGNAL 1 21 By similarity.
FT CHAIN 22 353 Outer membrane protein P5.
FT DISULFID 326 338 By similarity.
FT DOMAIN 272 316 OmpA-like.
SQ SEQUENCE 353 AA; 37743 MW; 64ACB3E7BFF96B39 CRC64;

Query Match 68.3%; Score 76.5; DB 1; Length 353;
Best Local Similarity 77.3%; Pred. NO. 0.0082;
Matches 17; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 1 RSDYKLYNKSSNSTLKLGE 22
DQ |||||:|||||
DB 132 RSDYKLYNKSSNSTLKLGE 150

RESULT 2
Q939P1 PRELIMINARY; PRT; 168 AA.
AC Q939P1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pBM02.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=11359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P8-2-47;
RX MEDLINE=22613826; PubMed=1272765;
RA Sanchez C., Mayo B.;
RT "Sequence and analysis of pBM02, a novel RCR cryptic plasmid from
RL Lactococcus lactis subsp cremoris P8-2-47.";
EL Plasmid 49:118-129(2003).
DR EMBL; AY026767; AAK13013.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 168 AA; 19397 MW; DB1B4221F94B2326 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 168;
Best Local Similarity 73.3%; Pred. NO. 3.6;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 KLYNKSSNSTLKN 19
DQ |||||:|||||
DB 149 KRYNKTSNSGLRN 163

RESULT 3
TIG_BUCAI
ID TIG_BUCAI STANDARD; PRT; 442 AA.
AC P57546;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Trigger factor (TF).
GN Name=tig; OrderedLocusNames=BU474;
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077; DOI=10.1038/35024074;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: Involved in protein export. Acts as a chaperone by
maintaining the newly synthesized protein in an open conformation

```

```

CC (By similarity).
CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family. Tig subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AP001119; BAB13171.1; -.
DR HAMAP; MF_00303; -.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR008880; Trigger_C.
DR InterPro; IPR008881; Trigger_N.
DR Pfam; PF00254; FKBP_C; 1.
DR Pfam; PF05698; Trigger_C; 1.
DR Pfam; PF05697; Trigger_N; 1.
DR PROSITE; PS50059; FKBP_PPIASE; FALSE NEG.
KW Cell division; Chaperone; Complete proteome; Isomerase; Rotamase.
FT DOMAIN 163 248 PPIase, FKBP-type.
SQ SEQUENCE 442 AA; 53482 MW; FCC99508F52150C2 CRC64;

Query Match 47.3%; Score 53; DB 1; Length 442;
Best Local Similarity 56.2%; Pred. No. 29;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 KLYNKSSNSTLKNL 20
DQ |||||:|||||
DB 388 KLYNKKNLKNLTKMNI 403

RESULT 4
Q8STT1 PRELIMINARY; PRT; 350 AA.
AC Q8STT1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein ECU09_0790.
GN Name=ECU09_0790;
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarar F.,
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL590451; CAD27052.1; -.
DR GO; GO:0004114; F:3', 5'-cyclic-nucleotide phosphodiesterase a. . .; IEA.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003607; Met_phos_hydro.
DR InterPro; IPR002073; PDBase.
DR Pfam; PF00233; PDBase I; 1.
DR SMART; SM00471; Hdc; 1.
KW Hypothetical protein.
SQ SEQUENCE 350 AA; 39803 MW; CE820F52979851B9 CRC64;

Query Match 46.9%; Score 52.5; DB 2; Length 350;

```

```

Best Local Similarity 35.5%; Pred. No. 27;
Matches 11; Conservative 6; Mismatches 3; Indels 11; Gaps 1;

Qy 3 DYKLYNKNSSS-----NSTLKNLGE 22
   |||:|||||: |||:|
Db 263 DYRFNKNSTAFWAEVYVDDDDYNETLENISE 293
   |||:|||||: |||:|

RESULT 5
Q82CA4 PRELIMINARY; PRT; 322 AA.
ID Q82CA4
AC Q82CA4;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocuNames=SAV5450;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=42608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AP005043; BAC73162; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 322 AA; 33126 MW; 2047F05D57BF7617 CRC64;

Query Match 46.4%; Score 52; DB 2; Length 322;
Best Local Similarity 56.5%; Pred. No. 29;
Matches 13; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

Qy 2 SDYKLYNK--NSSSNSTLKNLGE 22
   ||| | |||:|||||:
Db 143 SEYKRVKGGSSTSVTKDLAE 165
   ||| | |||:|||||:

RESULT 6
Q832Y3 PRELIMINARY; PRT; 322 AA.
ID Q832Y3
AC Q832Y3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN OrderedLocuNames=EF2085;
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacilliales; Enterococcaceae; Enterococcus.
NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=42550857; PubMed=12663927; DOI=10.1126/science.1080613;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,

```

```
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Ferredoxin (TrEMBLrel. 27, Last annotation update)
GN OrderedLocusNames=PT00948;
OS Picrophilus torridus.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Picrophilaceae; Picrophilus.
OX NCBI_TaxID=82076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 9790 / ATCC 70027;
RX PubMed=15184674;
RA Fuetterer O., Angelov A., Liesegang H., Gottschalk G., Schleper C.,
RT Schepers B., Dock C., Antranikian G., Liebl W.;
RT "Genome sequence of Picrophilus torridus and its implications for life
RT around pH 0.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9091-9096 (2004).
CC -!- FUNCTION: Catalyzes the ferrous insertion into protoporphyrin IX
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).
CC -!- PATHWAY: Protoheme biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the ferrochelatase family.
DR EMBL; AE017261; AAT43533.1; -.
DR GO; GO:0004325; F:ferrochelatase activity; IEA.
DR InterPro; IPR001015; Ferrochelatase.
DR Pfam; PF00762; Ferrochelatase; 1.
DR ProDom; PD002792; Ferrochelatase; 1.
KW Complete proteome; Heme biosynthesis; Iron; Lyase;
KW Porphyrin biosynthesis.
SQ SEQUENCE 287 AA; 33856 MW; 57773D598089D554 CRC64;

Query Match 44.6%; Score 50; DB 2; Length 287;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 40 RKKVELYNGRSPNSDILKCIIE 61

RESULT 9
Q7RA98 PRELIMINARY; PRT; 369 AA.
AC Q7RA98;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY06604;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=1236865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
SQ SEQUENCE 418 AA; 47948 MW; 7CC87D9D728DE1BC CRC64;

Query Match 44.6%; Score 50; DB 2; Length 418;
Best Local Similarity 42.1%; Pred. No. 76;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSTLKNL 20
Db 252 NNYNFFNKNKSKINEYIKNI 270

RESULT 11
Q96264 PRELIMINARY; PRT; 244 AA.
AC Q96264;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ST1967.
GN OrderedLocusNames=ST1967;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
```

```

RX MEDLINE=2145156; PubMed=11572479;
RA Kwarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshiwaza T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Maeda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.,
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.",
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000988; BAB67062.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR InterPro; IPR006342; FkM.
DR InterPro; IPR000051; SAM_bind.
DR TrkRFAMs; TRG01444; fkbM_fam; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 244 AA; 28182 MW; D19DCB1DABD9CF3 CRC64;

Query Match 44.2%; Score 49.5; DB 2; Length 244;
Best Local Similarity 51.9%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 6; Indels 7; Gaps 1;

Qy 2 SDYKLYNKSSNS-----TLNKG 21
Db 211 SDYKDVVKLSKNGYKIRVHTLNKG 237

RESULT 12
Q6GUB7 PRELIMINARY; PRT; 351 AA.
ID Q6GUB7
AC Q6GUB7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE OmpA.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95120769;
RA Akridge H., Confer A.W., Dabo S.M.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ompA family.
DR EMBL; AY643795; AAT57677.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR002368; OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA_LIKE.
DR InterPro; IPR000498; OmpA_tmam.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA_membrane; 1.
DR PRINTS; PR01021; OMFADOMAIN.
DR PRINTS; PR01022; OUTRMMERANEA.
DR PRODOM; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OMPA; 1.
SQ SEQUENCE 351 AA; 37694 MW; BCD8F661104354C1 CRC64;

Query Match 43.8%; Score 49; DB 2; Length 351;
Best Local Similarity 55.6%; Pred. No. 87;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSSTLK 18
Db 136 RSDYKVDKEPADLSFLK 153

RESULT 13
Q9ALY8 PRELIMINARY; PRT; 484 AA.
ID Q9ALY8
AC Q9ALY8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

---

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2523/90;
RA Oldfield N.J., Millar L.A., Ketley J.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF334762; AAK08092.1; -.
DR InterPro; IPR011029; DEATH_like.
KW Hypothetical protein.
SQ SEQUENCE 484 AA; 57093 MW; 679216E902915034 CRC64;

Query Match 43.8%; Score 49; DB 2; Length 484;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 KLYNKSSNSSTLKN 19
Db 204 KAHNKNSSSKITFKN 218

RESULT 14
Q7RF20 PRELIMINARY; PRT; 502 AA.
ID Q7RF20
AC Q7RF20;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein (Fragment).
GN Name=PY04891;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Portea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., O.R.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White M.J.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001514; EAA16830.1; -.
DR InterPro; IPR007201; Rrm_2.
DR Pfam; PF04059; RRM_2; 1.
KW Hypothetical protein.
FT NON TER 502
SQ SEQUENCE 502 AA; 58790 MW; 51E90077E7BD30E9 CRC64;

Query Match 43.8%; Score 49; DB 2; Length 502;
Best Local Similarity 47.6%; Pred. No. 1.3e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSSTLKNLGE 22
Db 33 SPVLYNKKGNSNQLDKLNK 53

```

```

RESULT 15
Q22762          PRELIMINARY;          PRT;    150 AA.
AC Q22762;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein T25B6.3.
GN ORFNames=T25B6.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favello A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41546; AAC48218.1; -.
DR PIR; T28904; T28904.
DR WormPep; T25B6.3; CE05013.
KW Hypothetical protein.
SQ SEQUENCE 150 AA; 17900 MW; F4B2944789F8FA4D CRC64;

Query Match          43.3%; Score 48.5; DB 2; Length 150;
Best Local Similarity 46.2%; Pred. No. 40;
Matches 12; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

Qy 1 RSDYKLYNK-----NSSNSTLKN 19
   :||: ||||| | || |||
Db 90 QSDHLYNKYFFHLGNFLSCTVKN 115

RESULT 16
QBILC9          PRELIMINARY;          PRT;    5561 AA.
AC Q8ILC9;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF14_0315;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

```

```

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014821; AAN36928.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; Znf_PHD.
DR SMART; SM00249; PHD; 4.
KW Hypothetical protein.
SQ SEQUENCE 5561 AA; 658527 MW; 98BF06CAC3C2B52B CRC64;

Query Match          43.3%; Score 48.5; DB 2; Length 5561;
Best Local Similarity 33.3%; Pred. No. 2.2e+03;
Matches 10; Conservative 6; Mismatches 3; Indels 11; Gaps 1;

Qy 4 YKLYNKSSS-----NSTLKNLGE 22
   |||: ||| | | | | | | | |
Db 3543 YKIYKKNKTSYKDDLNENIKDKSTIKNISD 3572

RESULT 17
Q7RHN4          PRELIMINARY;          PRT;    245 AA.
AC Q7RHN4;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Hypothetical protein (Fragment).
GN Name=PV03950;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmiller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001178; EAA15745.1; -.
KW Hypothetical protein.
FT NON TER 245
SQ SEQUENCE 245 AA; 28342 MW; 810043E5DFE8676F CRC64;

Query Match          42.9%; Score 48; DB 2; Length 245;
Best Local Similarity 52.9%; Pred. No. 82;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSNSTLKN 19
   :||| ||| | | | | |
Db 139 NYKLTNKNKEKNNNKN 155

RESULT 18
Q9CUF6          PRELIMINARY;          PRT;    384 AA.
ID Q9CUF6
AC Q9CUF6;

```



DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched  
 DE library, clone:4930577N17 product:hypothetical protein, full insert  
 DE sequence. (Fragment).  
 DN Name=4930577N17Rik;  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RL "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RL "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 403:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX The FANTOM Consortium;  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RL "Analysis of the mouse transcriptome based on functional annotation of  
 RL 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20493374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RL "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RL prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sudi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RL "RIKEN integrated sequence analysis (RISA) system-384-format  
 RL sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the ENBL/GenBank/DBJ databases.  
 DR EMBL; AK016297; BAB30183.1; -;  
 DR MGD; MGI:1914996; 4930577N17Rik.

KW Hypothetical protein.  
 FT NON-TER  
 SQ SEQUENCE 384 AA; 41209 MW; E9856A599F28262 CRC64;  
 Query Match 42.9%; Score 48; DB 2; Length 384;  
 Best Local Similarity 58.8%; Pred. No. 1.4e+02;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 Qy 4 YKLYNKSSNSTLKNL 20  
 Db 347 YATYQFSRKNTLKNL 363  
 RESULT 19  
 Q74BR0  
 ID Q74BR0 PRELIMINARY; PRT; 409 AA.  
 AC Q74BR0;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DE Hypothetical protein.  
 GN ORFNames=GSU1981;  
 OS Geobacter sulfurreducens.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
 OC Geobacteraceae; Geobacter.  
 OX NCBI\_TaxID=35554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCA / ATCC 51573;  
 RX PubMed=14671304; DOI=10.1126/science.1088727;  
 RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,  
 RA Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,  
 RA Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,  
 RA Winn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,  
 RA Davidse T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,  
 RA Weidman J.P., Khouz H.M., Feldblyum T.V., Utterback T.R.,  
 RA Van Aken S.E., Lovley D.R., Fraser C.M.;  
 RL "Genome of Geobacter sulfurreducens: metal reduction in subsurface  
 RL environments.";  
 RL Science 302:1967-1969(2003).  
 DR EMBL; AE017180; AAR35357.1; -;  
 DR TIGR; GSU1981; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 409 AA; 46361 MW; 27EABC4A3A7D6BEF CRC64;  
 Query Match 42.9%; Score 48; DB 2; Length 409;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 Qy 3 DYKLYNKSSNSTLKNL 20  
 Db 79 DYRYARNRSDDTHNL 96  
 RESULT 20  
 AAR35357  
 ID AAR35357 PRELIMINARY; PRT; 409 AA.  
 AC AAR35357;  
 DT 02-MAR-2004 (TREMBlrel. 27, Created)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DE Hypothetical protein.  
 GN GSU1981.  
 OS Geobacter sulfurreducens.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
 OC Geobacteraceae; Geobacter.  
 OX NCBI\_TaxID=35554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCA / ATCC 51573;  
 RX PubMed=14671304;  
 RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,  
 RA Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,

|           |  |
|-----------|--|
| RA        | Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,      |
| RA        | Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,       |
| RA        | Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A., |
| RA        | Weidman J., Khouri H.M., Feldblyum T.V., Utterback T.R.,               |
| RA        | Van Aken S.E., Lovley D.R., Fraser C.M.;                               |
| RT        | "Genome of Geobacter sulfurreducens: metal reduction in subsurface     |
| RT        | environments.";  |
| RL        | Science 302:1967-1969(2003).   |
| DR        | EMBL; AE017214; AAR35357.1; -.   |
| DR        | TIGR; GSU1981; -.  |
| KW        | Hypothetical protein.  |
| SQ        | SEQUENCE 409 AA; 43361 MW; 27EABC4A3A7D6BEF CRC64;                     |
|           | Query Match 42.9%; Score 48; DB 2; Length 409;                         |
|           | Best Local Similarity 50.0%; Pred. No. 1.5e+02;                        |
|           | Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;             |
| Qy        | 3 DYKLYNKVSSNSTLKNL 20   |
|           | :   :   :   :   :  |
| Db        | 79 DYRYAARNSRSDTTNHL 96  |
| RESULT 21 |  |
| Q8ZJD9    | PRELIMINARY; PRT; 706 AA.  |
| ID        | Q8ZJD9 Q74V13; Q7CFV4;   |
| AC        | 01-MAR-2002 (TrEMBLrel. 20, Created)                                   |
| DT        | 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)                      |
| DT        | 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)                      |
| DT        | 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)                    |
| DE        | Putative membrane protein (Hypothetical protein y3955).                |
| GN        | OrderedLocusNames=yP0173, YP00174, y3955;                              |
| OS        | Yersinia pestis.   |
| OC        | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;      |
| OC        | Enterobacteriaceae; Yersinia.  |
| OX        | NCBI_TaxID=632;  |
| RP        | [1]  |
| RP        | SEQUENCE FROM N.A.   |
| RC        | STRAIN=CO-92 / Biovar Orientalis;                                      |
| RC        | MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;               |
| RX        | Parkhill J., Wren B.W., Thomson N.R., Tittball R.W., Holden M.T.G.,    |
| RA        | Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,    |
| RA        | Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,   |
| RA        | Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,         |
| RA        | Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,         |
| RA        | Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,      |
| RA        | Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;       |
| RT        | "Genome sequence of Yersinia pestis, the causative agent of plague.";  |
| RL        | Nature 413:523-527(2001).  |
| RP        | [2]  |
| RP        | SEQUENCE FROM N.A.   |
| RC        | STRAIN=KIM5 / Biovar Mediaevalis;                                      |
| RX        | MEDLINE=22137863; PubMed=12142430;                                     |
| RA        | Deng W., Burland P., Plunkett G. III, Boutin A., Mayhew G.F., Liss P., |
| RA        | Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,                 |
| RA        | Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,              |
| RA        | Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., |
| RA        | Perry R.D.;  |
| RT        | "Genome sequence of Yersinia pestis KIM.";                             |
| RL        | J. Bacteriol. 184:4601-4611(2002).                                     |
| RP        | [3]  |
| RP        | SEQUENCE FROM N.A.   |
| RC        | STRAIN=91001 / Biovar Mediaevalis;                                     |
| RC        | Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D., |
| RA        | Han Y., Fang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,     |
| RA        | Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,      |
| RA        | Yang R.;   |
| RL        | Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.                |
| DR        | EMBL; AJ414141; CAC89036.1; -.   |
| DR        | EMBL; AE014000; AAM87499.1; -.   |
| DR        | EMBL; AE017127; AAS60450.1; -.   |
| DR        | PIR; AB0022; AB0022.   |
| DR        | InterPro; IPR009058; WING hlx DNA bnd.                                 |
| DR        | DUF893.  |

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196066; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazei R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.F., Agbayani A., An H.J., Andrews-Franks C., Baldwin D., Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouch J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A., Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A., George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R., Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J., Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C., Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RA "Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R., Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M., Ashburner M., Celniker S.E.;  
 RA "The transposable elements of the *Drosophila melanogaster* euchromatic genome: a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochman S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;  
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY119508; AM50162.1; -;  
 DR EMBL: AE003736; AAF55912.3; -;  
 DR MEROPS: M01.013; -;  
 DR FLYBase: FBgn0051233; CG31233;  
 DR GO: GO:0004179; F:membrane alanyl aminopeptidase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR001930; Peptidase M1.  
 DR Pfam: PF01433; Peptidase M1; 1.  
 DR PRINTS: PR00756; ALADIPPTASE.  
 SQ SEQUENCE 952 AA; 106857 MW; A42AB7D83C78812C CRC64;  
 Query Match 42.9%; Score 48; DB 2; Length 952;  
 Best Local Similarity 45.0%; Pred. No. 3.7e+02;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Qy 1 RSDYKLYNKNSSNSLTKNL 20  
 Db 330 RBYLLYNTENSTTQTNI 349  
 RESULT 24  
 Q8T148 PRELIMINARY; PRT; 1108 AA.  
 ID Q8T148  
 AC Q8T148;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Similar to Dictyostelium discoideum (Slime mold). Homeobox-containing protein.  
 DE Dictyostelium discoideum (Slime mold).  
 OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 OC NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RX MEDLINE=22092622; PubMed=12097910;  
 RA Gloeckner G., Eichinger L., Szatranski K., Pachebat J., Dear P., Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K., Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
 RA "Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*.";  
 RL Nature 418:79-85(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Baumgart C.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC117070; AM09320.2; -;  
 DR GO: GO:0005634; C:nucleus; IEA.  
 KW Nuclear protein.  
 SQ SEQUENCE 1108 AA; 124003 MW; 167437F1518F1BEE CRC64;  
 Query Match 42.9%; Score 48; DB 2; Length 1108;  
 Best Local Similarity 66.7%; Pred. No. 4.4e+02;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 5 KLYNKNSSNSLTKN 19  
 Db 80 KLYNKNSSNSPIIN 94  
 RESULT 25  
 Q6LEX8 PRELIMINARY; PRT; 6761 AA.  
 ID Q6LEX8  
 AC Q6LEX8;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Putative SET-domain protein (Set-domain protein, putative).  
 GN ORFNames=WAL6P1.131, PPF1440w;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,  
 RA Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James D., Johnson D., Kerhornou A., Knight A.,  
 RA Kontfortov B., Keyes S., Larke N., Lawson D., Lennard N., Line A.,  
 RA Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,  
 RA Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M-A.,  
 RA Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,  
 RA Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,  
 RA Tivey A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A.,  
 RA Newbold C., Barrell B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,  
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.G;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382401; CAG25109.1; -;  
 DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
 DR InterPro; IPR001487; Bromodomain.  
 DR InterPro; IPR000595; cNMP binding.  
 DR InterPro; IPR010983; EF Hand like.  
 DR InterPro; IPR011009; Kinase Like.  
 DR InterPro; IPR003616; PostSET.  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR001965; ZnF\_PHD.  
 DR Pfam; PF00439; Bromodomain; 1.  
 DR Pfam; PF00628; PHD; 1.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00297; BRON; 1.  
 DR SMART; SM00249; PHD; 4.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; UNKNOWN\_1.  
 DR PROSITE; PS50014; BROMODOMAIN 2; 1.  
 DR PROSITE; PS50868; POST\_SET; 1.  
 DR PROSITE; PS50280; SET; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; UNKNOWN\_1.  
 DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
 SQ SEQUENCE 6761 AA; 797029 MW; 6FD6CA835134F303 CRC64;  
  
 Query Match 42.9%; Score 48; DB 2; Length 6761;  
 Best Local Similarity 36.4%; Pred. No. 3.2e+03;  
 Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
  
 Qy 1 RSDYKLYKNSSNSTLKNLGE 22  
 Db 5132 KDDHHYNNQNNHNSYLCIDIGK 5153  
  
 RESULT 26  
 ARLY\_CLOTE  
 ID ARLY\_CLOTE STANDARD; PRT; 438 AA.  
 AC P59616;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL).  
 GN Name=argh; OrderedLocusNames=CTC00562;  
 OS Clostridium tetani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 NCBI\_TaxID=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88;  
 RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;  
 RA Brueggemann H., Baumer S., Fricke W.F., Wierer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.;  
 RT "The genome sequence of Clostridium tetani, the causative agent of  
 RT tetanus disease."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 CC -|- CATALYTIC ACTIVITY: N-(L-arginino)succinate = fumarate + L-  
 CC arginine.  
 CC -|- PATHWAY: Arginine biosynthesis; eighth (last) step.  
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (probable).

DE Putative SET-domain protein (SET-domain protein, putative).  
 GN MAL6P1.131.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium;  
 NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RC PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,  
 RA Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James D., Johnson D., Kerhornou A., Knight A.,  
 RA Kontfortov B., Keyes S., Larke N., Lawson D., Lennard N., Line A.,  
 RA Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,  
 RA Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M-A.,  
 RA Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,  
 RA Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,  
 RA Tivey A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A.,  
 RA Newbold C., Barrell B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,  
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.G;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382401; CAG25109.1; -;  
 SQ SEQUENCE 6761 AA; 797029 MW; 6FD6CA835134F303 CRC64;  
  
 Query Match 42.9%; Score 48; DB 2; Length 6761;  
 Best Local Similarity 36.4%; Pred. No. 3.2e+03;  
 Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
  
 Qy 1 RSDYKLYKNSSNSTLKNLGE 22  
 Db 5132 KDDHHYNNQNNHNSYLCIDIGK 5153  
  
 RESULT 27  
 ARLY\_CLOTE  
 ID ARLY\_CLOTE STANDARD; PRT; 438 AA.  
 AC P59616;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL).  
 GN Name=argh; OrderedLocusNames=CTC00562;  
 OS Clostridium tetani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 NCBI\_TaxID=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88;  
 RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;  
 RA Brueggemann H., Baumer S., Fricke W.F., Wierer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.;  
 RT "The genome sequence of Clostridium tetani, the causative agent of  
 RT tetanus disease."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 CC -|- CATALYTIC ACTIVITY: N-(L-arginino)succinate = fumarate + L-  
 CC arginine.  
 CC -|- PATHWAY: Arginine biosynthesis; eighth (last) step.  
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (probable).

```

CC CC -1- SIMILARITY: Belongs to the lyase 1 family. Argininosuccinate lyase
CC CC subfamily.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC CC or send an email to license@ebi.ac.uk).
CC CC -----
CC DR EMBL; AE015937; AA035181.1; -.
CC DR HSSP; P24058; IDCN.
CC DR HAMAP; MF_00006; -.
CC DR InterPro; IPR009049; argH.
CC DR InterPro; IPR003031; D_cryotallin.
CC DR InterPro; IPR000362; Fumarate lyase.
CC DR InterPro; IPR008948; L-Aspartase-like.
CC DR Pfam; PF00206; Lyase_1; 1.
CC DR PRINTS; PR00145; DCRYSTALLIN.
CC DR PRINTS; PR00149; FUMRATLYASE.
CC DR TIGRFAMs; TIGR00838; argH; 1.
CC DR PROSITE; PS00163; FUMARATE LYASES; 1.
CC DR Arginine biosynthesis; Complete proteome; Lyase.
CC KW Arginine biosynthesis; Complete proteome; Lyase.
CC SQ SEQUENCE 438 AA; 50296 MW; 97CCEDIE27807C53 CRC64;

Query Match 42.4%; Score 47.5; DB 1; Length 438;
Best Local Similarity 41.4%; Pred. No. 1.9e+02;
Matches 12; Conservative 3; Mismatches 5; Indels 9; Gaps 1;

Qy 3 DYKLYNKNS-----SNSLKNLGE 22
Db 115 DFLYAKNLSLKIESIETLQNTKLNLA 143

RESULT 28
Q7RLW6 PRELIMINARY; PRT; 595 AA.
AC Q7RLW6;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA-directed RNA polymerase III.
GN Name=PY02422;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Perteza M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.B.,
RA Shalom S.J., van Aken S.E., Riedmuller S.B., Feldbush T.V.,
RA Cho J.K., Quackenbush J., Sedegh M., Shoaibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000665; EAA21865.1; -.
CC GO; GO:0003899; F-DNA-directed RNA polymerase activity; IEA.
CC KW DNA-directed RNA polymerase.
CC SQ SEQUENCE 595 AA; 71812 MW; 0C546431C07F5663 CRC64;

Query Match 42.4%; Score 47.5; DB 2; Length 595;

```

```

Best Local Similarity 57.9%; Pred. No. 2.6e+02;
Matches 11; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 1 RSDYKLYNKSSNSSTLKN 19
Db 456 KDDYLL-NKNSNNYNNILKN 473

RESULT 29
HRL2 LACMU STANDARD; PRT; 200 AA.
ID HRL2 LACMU
AC P22756;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Haemorrhagic factor II (EC 3.4.24.-) (LHF-II).
OS Lachesis muta muta (Bushmaster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Lachesis.
OX NCBI_TaxID=8753;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=91224327; PubMed=2026257;
RA Sanchez E.F., Diniz C.R., Richardson M.;
RT "The complete amino acid sequence of the haemorrhagic factor LHP-II, a
RT metalloproteinase isolated from the venom of the bushmaster snake
RT (Lachesis muta muta).";
RL FEBS Lett. 282:178-182(1991).
CC -1- FUNCTION: This protein is a zinc protease from snake venom that
CC acts in hemorrhage.
CC -1- COFACTOR: Binds 1 zinc ion and 2 calcium ions per subunit.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: Belongs to peptidase family M12B.
DR PIR; S1511; S1511.
DR HSP; P15167; IATL.
DR MEROPS; M12.162; -.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01421; Reprolysin_1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Calcium-binding; Direct protein sequencing; Hydrolase; Metal-binding;
KW Metalloprotease; Zinc.
FT METAL 7 7 Calcium (By similarity).
FT METAL 91 91 Calcium (By similarity).
FT METAL 140 140 Zinc (catalytic) (By similarity).
FT ACT_SITE 141 141 By similarity.
FT METAL 144 144 Zinc (catalytic) (By similarity).
FT METAL 150 150 Zinc (catalytic) (By similarity).
FT METAL 198 198 Calcium (By similarity).
FT VARIANT 134 134 I -> T.
SQ SEQUENCE 200 AA; 22596 MW; 934954B56785DB13 CRC64;

Query Match 42.0%; Score 47; DB 1; Length 200;
Best Local Similarity 47.6%; Pred. No. 92;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSSTLKNLGE 22
Db 58 SNOQLINVQSAANDTLKTFGE 78

RESULT 30
Q7K6A7 PRELIMINARY; PRT; 229 AA.
ID Q7K6A7
AC Q7K6A7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Cyclin2 related protein, putative.
GN Name=PFE0920c;

```

OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerkhoun A.,  
 RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sultston J.E., Craig A., Newbold C., Barrell B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the cyclin family.  
 DR EMBL; AL293353; CAD51548.1; -;  
 DR InterPro; IPR011028; Cyclin like.  
 DR Pfam; PF00134; Cyclin\_N.  
 KW Cyclin.  
 SQ SEQUENCE 229 AA; 26846 MW; D34490978209D930 CRC64;  
 Query Match 42.0%; Score 47; DB 2; Length 229;  
 Best Local Similarity 52.9%; Pred. No. 1.1e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 RSDYKLYNKSSNSTL 17  
 Db 180 KKPYNLFNYKSSNSTL 196  
 RESULT 31  
 ID Q8T357 PRELIMINARY; PRT; 229 AA.  
 AC Q8T357  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative cyclin 3.  
 GN Name=cyc-3;  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22890188; PubMed=12869562;  
 RA Merckx A.F., Le Roch K., Nivez M.P., Dorin D., Alano P.,  
 RA Gutierrez G.J., Nebreda A.R., Goldring D., Whittle C., Patterson S.,  
 RA Chakrabarti D., Doerig C.;  
 RT "Identification and initial characterization of three novel cyclin-  
 RT related proteins of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL J. Biol. Chem. 278:39839-39850(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Doerig C.O.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: Belongs to the cyclin family.  
 DR EMBL; AJ416697; CAC95051.1; -;  
 DR InterPro; IPR011028; Cyclin like.  
 DR Pfam; PF00134; Cyclin\_N.  
 KW Cyclin.  
 SQ SEQUENCE 229 AA; 26846 MW; D34490978209D930 CRC64;  
 Query Match 42.0%; Score 47; DB 2; Length 229;  
 Best Local Similarity 52.9%; Pred. No. 1.1e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 RSDYKLYNKSSNSTL 17  
 Db 180 KKPYNLFNYKSSNSTL 196  
 RESULT 32  
 ID Q6DKA7 PRELIMINARY; PRT; 246 AA.  
 AC Q6DKA7  
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnurch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative."  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RA Klein S., Strausberg R.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 KW EMBL; BC074453; AAH74453.1; -;  
 RT Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 246 AA; 27636 MW; B67B0E0988C2E1A7 CRC64;

```

Query Match      42.0%; Score 47; DB 2; Length 246;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 YNKNSSNSTLKNLGE 22
   |||||:|:|:|
Db 81 YNNSNSSSTTTQNRAE 96

RESULT 33
O50827 PRELIMINARY; PRT; 308 AA.
AC O50827;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein BBK23.
GN OrderedLocusNames=BBK23;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp36.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Winn M.L., Dougherty B.A., Tomb J.-P., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi".
RL Nature 390:580-586(1997).
DR EMBL; AE000788; AAC66174.1; -.
DR TIGR; A70254; A70254.
DR InterPro; IPR003459; Borrelia Unk.
DR Pfam; PF02414; Borrelia orfA_1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 308 AA; 36494 MW; E8233C4660C5A015 CRC64;

Query Match      42.0%; Score 47; DB 2; Length 308;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 6 LYNKNSNSTLKNL 20
   :|||:|:|:|
Db 183 IYNKNSIENSFLKKI 197

RESULT 34
Q6L273 PRELIMINARY; PRT; 426 AA.
AC Q6L273;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=MMP0756;
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / LL;
RA Hendrickson E.L., Kaul R., Zhou Y., Boyee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeiri A., Rouse G.,
RA Saenphimmachak C., Soell D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT methanogen Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957221; CAF30312.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 426 AA; 50079 MW; 2E07C24D217F6AC6 CRC64;

Query Match      42.0%; Score 47; DB 2; Length 426;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLGE 22
   :|||:|:|:|
Db 201 KSTIKYLLKNSSLSNLSNLRINE 222

RESULT 35
CAF30312 PRELIMINARY; PRT; 426 AA.
AC CAF30312;
DT 03-MAR-2004 (TrEMBLrel. 27, Created)
DT 03-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN MMP0756.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / LL;
RA Hendrickson E.L., Kaul R., Zhou Y., Boyee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeiri A., Rouse G.,
RA Saenphimmachak C., Soell D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT methanogen Methanococcus maripaludis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX957221; CAF30312.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 426 AA; 50079 MW; 2E07C24D217F6AC6 CRC64;

Query Match      42.0%; Score 47; DB 2; Length 426;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLGE 22
   :|||:|:|:|
Db 201 KSTIKYLLKNSSLSNLSNLRINE 222

RESULT 36
Q8W2N4 PRELIMINARY; PRT; 510 AA.
AC Q8W2N4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450-dependent fatty acid hydroxylase.
GN Names=CYP94A6;
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.

```





Db 762 YEYKNGAPNSTV 775  
|:||||:|:

RESULT 40  
Q7RCN2  
ID Q7RCN2 PRELIMINARY; PRT; 1078 AA.  
AC Q7RCN2  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Adapter-related protein complex 1 gamma 1 subunit.  
GN Name=PY05746;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=73239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17XNL;  
RX PubMed=12368865;  
RA Cariton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,  
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
RA Carucci D.J.;  
RT "Genome sequence and comparative analysis of the model rodent malaria  
RT parasite Plasmodium yoelii yoelii.";  
RL Nature 419:512-519(2002).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AABL01001863; EAA17814.1; -;  
DR GO; GO:0030130; C:clathrin coat of trans-Golgi network vesicle; IEA.  
DR GO; GO:0006886; P:intracellular protein transport; IEA.  
DR GO; GO:0006461; P:protein complex assembly; IEA.  
DR InterPro; IPR008152; A/G adapt C.  
DR InterPro; IPR002553; Adaptin\_N.  
DR InterPro; IPR008153; Gamma\_adaptin\_C.  
DR Pfam; PF01602; Adaptin\_N; I.  
DR Pfam; PF02883; Alpha\_adaptinC2; I.  
DR ProDom; PD021457; Gamma\_adaptin\_C; I.  
DR PROSITE; PS50180; GAE; I.  
SQ SEQUENCE 1078 AA; 122919 MW; 6AC61A9B32CC649B CRC64;

Query Match 42.0%; Score 47; DB 2; Length 1078;  
Best Local Similarity 62.5%; Pred. No. 5.9e+02;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 6 LYNNSSNSTLKNIG 21  
|||||:|:  
Db 622 LYNNSGNSRSGVLG 637

Search completed: November 24, 2004, 09:28:53  
Job time : 105.152 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:07:49 ; Search time 82.8152 Seconds  
(without alignments)  
82.302 Million cell updates/sec

Title: US-09-719-379A-4  
Perfect score: 98  
Sequence: 1 RSDYKLYNKSSITLKDLGE 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 98    | 100.0       | 19     | 3     | AAY79978 Non-typea |
| 2          | 98    | 100.0       | 19     | 3     | AAY79990 Non-typea |
| 3          | 98    | 100.0       | 19     | 3     | AAY79984 Non-typea |
| 4          | 98    | 100.0       | 19     | 4     | AAB47442 Lb1(f) co |
| 5          | 98    | 100.0       | 28     | 4     | AAB47446 Entire 3r |
| 6          | 98    | 100.0       | 353    | 5     | ABG80420 Haemophil |
| 7          | 87.5  | 89.3        | 20     | 3     | AAY79979 Non-typea |
| 8          | 87    | 88.8        | 353    | 4     | AAB47447 MOMP P5.  |
| 9          | 82    | 83.7        | 21     | 3     | AAY79977 Non-typea |
| 10         | 82    | 83.7        | 21     | 3     | AAY79975 Non-typea |
| 11         | 81.5  | 83.2        | 22     | 3     | AAY79988 Non-typea |
| 12         | 81.5  | 83.2        | 22     | 3     | AAY79983 Non-typea |
| 13         | 81.5  | 83.2        | 22     | 3     | AAY79974 Non-typea |
| 14         | 81.5  | 83.2        | 22     | 4     | AAB47440 Lb1(f) co |
| 15         | 81.5  | 83.2        | 31     | 4     | AAB47444 Entire 3r |
| 16         | 81.5  | 83.2        | 464    | 3     | AAY79993 Plasmid L |
| 17         | 75.5  | 77.0        | 20     | 3     | AAY79980 Non-typea |
| 18         | 75    | 76.5        | 21     | 3     | AAY79976 Non-typea |
| 19         | 49    | 50.0        | 360    | 2     | Aaw98912 RPS2 prot |
| 20         | 49    | 50.0        | 683    | 2     | Aaw98915 IER prot  |
| 21         | 49    | 50.0        | 877    | 2     | Aaw98914 IER prote |
| 22         | 49    | 50.0        | 917    | 2     | AAR80132 Rps2 dise |
| 23         | 49    | 50.0        | 965    | 2     | AAR83440 A.thalian |
| 24         | 49    | 50.0        | 1054   | 2     | Aaw98913 CER prote |
| 25         | 48    | 49.0        | 365    | 6     | ABU17574 Protein e |

|    |      |      |      |   |          |                     |
|----|------|------|------|---|----------|---------------------|
| 26 | 48   | 49.0 | 536  | 6 | ABU17880 | Abu17880 Protein e  |
| 27 | 47   | 48.0 | 200  | 4 | ABB61039 | Abb61039 Drosophil  |
| 28 | 46   | 46.9 | 648  | 6 | ABU24397 | Abu24397 Protein e  |
| 29 | 46   | 46.9 | 1829 | 2 | AAW29322 | AAW29322 DNA polym  |
| 30 | 46   | 46.9 | 1829 | 8 | ADN59952 | ADN59952 Thermococ  |
| 31 | 45   | 45.9 | 143  | 4 | AAO00088 | AAO00088 Human pol  |
| 32 | 45   | 45.9 | 311  | 5 | ABP30404 | ABP30404 Streptoco  |
| 33 | 45   | 45.9 | 315  | 5 | ABP27295 | ABP27295 Streptoco  |
| 34 | 45   | 45.9 | 624  | 5 | ABP73960 | ABP73960 Candida a  |
| 35 | 45   | 45.9 | 1017 | 8 | ADN73899 | ADN73899 Thale cre  |
| 36 | 44.5 | 45.4 | 733  | 6 | ADA34312 | Ada34312 Acinetoba  |
| 37 | 44   | 44.9 | 96   | 5 | ADK34379 | Adk34379 Novel hum  |
| 38 | 44   | 44.9 | 140  | 3 | AAG22869 | Aag22869 Arabinidop |
| 39 | 44   | 44.9 | 550  | 5 | AAM50854 | Aam50854 Comamonas  |
| 40 | 44   | 44.9 | 1058 | 5 | ABR52727 | ABR52727 Protein s  |
| 41 | 44   | 44.9 | 1058 | 7 | ADK61962 | Adk61962 Disease t  |
| 42 | 44   | 44.9 | 1216 | 6 | ABR53002 | ABR53002 Protein s  |
| 43 | 44   | 44.9 | 1216 | 7 | ADK62734 | Adk62734 Disease t  |
| 44 | 43   | 43.9 | 19   | 3 | AAY79962 | Aay79962 Non-typea  |
| 45 | 43   | 43.9 | 58   | 8 | ADN63374 | Adn63374 Tobamovir  |

ALIGNMENTS

RESULT 1  
AAY79978

ID AAY79978 standard; peptide; 19 AA.

XX AC AAY79978;

XX DT 15-MAY-2000 (first entry)

XX DE Non-typeable H. influenzae group 2 Lb1(f) peptide N1236WEE.

XX KW Vaccine; non-typeable Haemophilus influenzae; ntl; infection;  
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.

XX OS Haemophilus influenzae.

XX PN WO9964067-A2.

XX PD 16-DEC-1999.

XX PF 28-MAY-1999; 99WO-US011980.

XX PR 11-JUN-1998; 98GB-00012613.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
(OHIS ) UNIV OHIO STATE RES FOUND.

XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX XX WPI; 2000-116457/10.

XX DR Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
Haemophilus influenza.

XX PS Example 1; Page 30; 68pp; English.

XX CC The present invention describes antigenic P5-like fimbria subunit  
peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
Haemophilus influenzae strains. The peptides are used for diagnosis,  
prevention, and treatment of Haemophilus influenzae infections, such as  
otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
infection. The peptides may also be used in vaccines against H.  
influenzae. Antibodies and probes from the present invention can be used  
for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
the present invention

SQ Sequence 19 AA;

Query Match 100.0%; Score 98; DB 3; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19  
 |||||  
 Db 1 RSDYKLYNKSSTLKDLGE 19  
 |||||

RESULT 2  
 AAY79990  
 ID AAY79990 standard; peptide; 19 AA.  
 XX  
 AC AAY79990;  
 XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Non-typeable H. influenzae 183NP Group 2b type peptide.  
 XX  
 KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO9964067-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 DR WPI; 2000-116457/10.  
 XX  
 PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 Haemophilus influenza.  
 XX  
 PS Example 1; Page 30; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbria subunit  
 peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 Haemophilus influenzae strains. The peptides are used for diagnosis,  
 prevention, and treatment of Haemophilus influenzae infections, such as  
 otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 infection. The peptides may also be used in vaccines against H.  
 influenzae. Antibodies and probes from the present invention can be used  
 for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 AA291201 to AA291252, represent sequences used in the exemplification of  
 the present invention

SQ Sequence 19 AA;

Query Match 100.0%; Score 98; DB 3; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19  
 |||||  
 Db 1 RSDYKLYNKSSTLKDLGE 19  
 |||||

RESULT 3  
 AAY79984  
 ID AAY79984 standard; peptide; 19 AA.

XX AAY79984;  
 AC  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Non-typeable H. influenzae group 2b Lb1(f) peptide NTHI-183.  
 XX  
 KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO9964067-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 DR WPI; 2000-116457/10.  
 XX  
 PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 Haemophilus influenza.  
 XX  
 PS Example 1; Page 30; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbria subunit  
 peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 Haemophilus influenzae strains. The peptides are used for diagnosis,  
 prevention, and treatment of Haemophilus influenzae infections, such as  
 otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 infection. The peptides may also be used in vaccines against H.  
 influenzae. Antibodies and probes from the present invention can be used  
 for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 AA291201 to AA291252, represent sequences used in the exemplification of  
 the present invention

SQ Sequence 19 AA;

Query Match 100.0%; Score 98; DB 3; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19  
 |||||  
 Db 1 RSDYKLYNKSSTLKDLGE 19  
 |||||

RESULT 4  
 AAB47442  
 ID AAB47442 standard; peptide; 19 AA.  
 XX  
 AC AAB47442;  
 XX  
 DT 31-OCT-2001 (first entry)  
 XX  
 DE Lb1(f) cont. peptide from strain nH1-183NP (Group 2b type).  
 XX  
 KW surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; nH1; Lb1(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX

PN WO200161013-A1.  
 PD 23-AUG-2001.  
 XX 13-FEB-2001; 2001WO-EP001556.  
 XX 15-FEB-2000; 2000GB-00003502.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Berthet FJ, Denoel P, Poolman J, Thonnard J;  
 XX WPI; 2001-522599/57.  
 XX Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.  
 XX Claim 1; Page 26; 29pp; English.  
 XX The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (ntHi). Each of these  
 CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain ntHill28, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX Sequence 19 AA;  
 SQ  
 Query Match 100.0%; Score 98; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKLYNKNSSTLKDLGE 19  
 DB 1 RSDYKLYNKNSSTLKDLGE 19  
 RESULT 5  
 AAB47446  
 ID AAB47446 standard; peptide; 28 AA.  
 XX AAB47446;  
 XX 31-OCT-2001 (first entry)  
 DT Entire 3rd loop from strain ntHi-183NP (Group 2b type).  
 DE surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; ntHi; LBI(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.  
 XX Haemophilus influenzae.  
 OS  
 XX WO200161013-A1.  
 XX 23-AUG-2001.  
 XX 13-FEB-2001; 2001WO-EP001556.  
 XX 15-FEB-2000; 2000GB-00003502.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX WPI; 2001-522599/57.  
 XX Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.  
 XX Claim 2; Page 26; 29pp; English.  
 XX The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (ntHi). Each of these  
 CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain ntHill28, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX Sequence 28 AA;  
 SQ  
 Query Match 100.0%; Score 98; DB 4; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKLYNKNSSTLKDLGE 19  
 DB 3 RSDYKLYNKNSSTLKDLGE 21  
 RESULT 6  
 AAB80420  
 ID AAB80420 standard; protein; 353 AA.  
 XX AAB80420;  
 XX 29-NOV-2002 (first entry)  
 DT Haemophilus influenzae P5 protein.  
 DE Hyperblebbing; Gram-negative bacterium; genetically modified; Tol gene;  
 KW peptidoglycan-binding; peptidoglycan-associated site; outer-membrane;  
 KW bacterial infection; vesicle-shedding; Bleb; filter sterilised;  
 KW detergent; deoxycholate; homogeneity; antibacterial; vaccine; HiRD; P5;  
 KW P6; PCP.  
 XX Haemophilus influenzae.  
 OS  
 XX WO200262378-A2.  
 XX 15-AUG-2002.  
 XX 08-FEB-2002; 2002WO-EP001361.  
 XX 08-FEB-2001; 2001GB-00003171.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Berthet FJ, Denoel P, Neyt CA, Poolman J, Thonnard J;  
 XX WPI; 2002-657509/70.  
 XX N-PSDB; ABB66193.  
 XX Novel hyperblebbing Gram-negative bacterium that has been genetically  
 PT modified by down-regulating expression of Tol genes, and/or attenuating  
 PT peptidoglycan-binding activity useful for treating bacterial infection.  
 XX Disclosure; Page 58; 71pp; English.

CC The present invention relates to a new hyperblebbing Gram-negative  
CC bacterium genetically modified by one or more processes selected from  
CC down-regulating expression of one or more Tol genes and attenuating the  
CC peptidoglycan-binding activity by mutation of one or more gene(s)  
CC encoding a protein comprising a peptidoglycan-associated site. The  
CC invention is useful in a method of treatment of the human or animal body.  
CC The invention is also useful for protecting an individual against a  
CC bacterial infection. The invention has improved outer-membrane vesicle-  
CC shedding properties. Bles are more easily made in higher yield from the  
CC invention, and are more homogeneous in nature and can be more readily  
CC filter sterilised. The bles can be made and harvested without the use of  
CC detergents such as deoxycholate, thus obviating chromatography  
CC purification and ultra centrifugation steps. Vesicles prepared from the  
CC invention have reduced particle size (allowing sterile filtration through  
CC 0.22 µm pores), increased batch homogeneity, and a superior yield. The  
CC present amino acid sequence represents a Haemophilus influenzae protein,  
CC as described in the invention  
XX  
SQ Sequence 353 AA;

Query Match 100.0%; Score 98; DB 5; Length 353;  
Best Local Similarity 100.0%; Pred. No. 5.6e-08; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0;

Qy 1 RSDYKLYNKNSTLKDGLGE 19  
Db 132 RSDYKLYNKNSTLKDGLGE 150  
|||||

RESULT 7  
AAV79979  
ID AAV79979 standard; peptide; 20 AA.  
AC AAV79979;

DT 15-MAY-2000 (first entry)

DE Non-typeable H. influenzae group 2 LB1(f) peptide N165NP.

KW Vaccine; non-typeable Haemophilus influenzae; nH; infection;  
KW Chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.

OS Haemophilus influenzae.

FN WO9964067-A2.

PD 16-DEC-1999.

PF 28-MAY-1999; 99WO-US011980.

PR 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.

PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
PT Haemophilus influenzae.

PS Example 1; Page 30; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
CC prevention, and treatment of Haemophilus influenzae infections, such as  
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
CC infection. The peptides may also be used in vaccines against H.  
CC influenzae. Antibodies and probes from the present invention can be used

CC for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and  
CC AA91201 to AA91252, represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 20 AA;

Query Match 89.3%; Score 87.5; DB 3; Length 20;  
Best Local Similarity 95.0%; Pred. No. 1e-07;  
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 RSDYKLYNKNSS-TLKDGLGE 19  
Db 1 RSDYKLYNKNSSNTLKDGLGE 20  
|||||

RESULT 8

AAV7447

ID AAB47447 standard; protein; 353 AA.

AC AAB47447;

XX 31-OCT-2001 (first entry)

XX MOMP P5.

DE surface exposed loop; major outer membrane protein P5; MOMP P5;  
KW non-typeable H. influenzae; nH; LB1(f) peptide; B cell epitope;  
KW otitis media; sinusitis; conjunctivitis;

XX lower respiratory tract infection.

OS Haemophilus influenzae.

FN Key Location/Qualifiers

FT Domain 38..57  
FT /label= Loop 1  
FT /note= "Extracellular domain"

FT Domain 89..100  
FT /label= Loop 2  
FT /note= "Extracellular domain"

FT Domain 136..150  
FT /label= Loop 3  
FT /note= "Extracellular domain"

FT Domain 181..204  
FT /label= Loop 4  
FT /note= "Extracellular domain"

XX WO200161013-A1.

XX 23-AUG-2001.

XX 13-FEB-2001; 2001WO-EP001556.

XX 15-FEB-2000; 2000GB-00003502.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX WPI; 2001-522599/57.

XX Recombinant bacterial outer membrane protein where one or more surface-  
PT exposed loops are modified is useful as a vaccine to prevent or treat  
PT Haemophilus influenzae infection or associated disease, e.g., otitis  
PT media and conjunctivitis.

PS Disclosure; Fig 1; 29pp; English.

XX This sequence represents the major outer membrane protein P5 of non-  
CC typeable H. influenzae. One or more surface exposed loops of this protein  
CC may be replaced with a modified peptide of the invention. Each of these  
CC peptides contain an LB1(f) peptide which is a 19 amino acid peptide  
CC derived from the sequence of MOMP P5 from strain nH1128, representing  
CC amino acids Arg117 to Gly135. This peptide represents the third exposed

CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX  
 SQ Sequence 353 AA;

Query Match 88.8%; Score 87; DB 4; Length 353;

Best Local Similarity 89.5%; Pred. No. 4.3e-06;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDLGE 19

Db 132 RSDYKLYNKSSTLKDLGE 150

RESULT 9

AAV79977

ID AAV79977 standard; peptide; 21 AA.

XX

AC AAV79977;

XX

DT 15-MAY-2000 (first entry)

XX

DE Non-typeable H. influenzae group 2 LB1 (f) peptide NTHI-506.

XX

KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1 (f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX

OS Haemophilus influenzae.

XX

PN WO9964067-A2.

XX

PD 16-DEC-1999.

XX

PF 28-MAY-1999; 99WO-US011980.

XX

PR 11-JUN-1998; 98GB-00012613.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PA (OHIS) UNIV OHIO STATE RES FOUND.

XX

PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX

DR WPI; 2000-116457/10.

XX

PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenzae.

XX

PS Example 1; Page 30; 68pp; English.

XX

CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1 (f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.

CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention

XX

SQ Sequence 21 AA;

Query Match 83.7%; Score 82; DB 3; Length 21;

Best Local Similarity 85.7%; Pred. No. 9.3e-07;

Matches 18; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

Qy 1 RSDYKLYNKS--STLKDLGE 19

Db 1 RSDYKLYNKSSTLKDLGE 21

RESULT 10

AAV79975

ID AAV79975 standard; peptide; 21 AA.

XX

AC AAV79975;

XX

DT 15-MAY-2000 (first entry)

XX

DE Non-typeable H. influenzae group 2 LB1 (f) peptide NTHI-492.

XX

KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1 (f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX

OS Haemophilus influenzae.

XX

PN WO9964067-A2.

XX

PD 16-DEC-1999.

XX

PF 28-MAY-1999; 99WO-US011980.

XX

PR 11-JUN-1998; 98GB-00012613.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PA (OHIS) UNIV OHIO STATE RES FOUND.

XX

PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX

DR WPI; 2000-116457/10.

XX

PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenzae.

XX

PS Example 1; Page 30; 68pp; English.

XX

CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1 (f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.

CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention

XX

SQ Sequence 21 AA;

Query Match 83.7%; Score 82; DB 3; Length 21;

Best Local Similarity 85.7%; Pred. No. 9.3e-07;

Matches 18; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

Qy 1 RSDYKLYNKS--STLKDLGE 19

Db 1 RSDYKLYNKSSTLKDLGE 21

RESULT 11

AAV79988

ID AAV79988 standard; peptide; 22 AA.

XX

AC AAV79988;

XX

DT 15-MAY-2000 (first entry)

XX

DE Non-typeable H. influenzae 1715MEE Group 2a type peptide.

XX

KW Vaccine; non-typeable *Haemophilus influenzae*; nH<sub>i</sub>; infection;  
 KW chimeric protein; *Haemophilus influenzae*; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS *Haemophilus influenzae*.  
 XX  
 PN WO9964067-A2.  
 XX  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI  
 XX WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT *Haemophilus influenzae*.  
 XX  
 XX Claim 3; Page 46; 68pp; English.  
 PS  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC *Haemophilus influenzae* strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of *Haemophilus influenzae* infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC *influenzae*. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. *influenzae* infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 22 AA;  
 SQ  
 Query Match 83.2%; Score 81.5; DB 3; Length 22;  
 Best Local Similarity 81.8%; Pred. No. 1.2e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 RSDYKLYNKNS---STLKDGLG 19  
 DB 1 RSDYKLYNKNSSSNSTLKNLGE 22  
 ||||| ||||| ||||| |||||  
 RESULT 12  
 AAY79983  
 ID AAY79983 standard; peptide; 22 AA.  
 AC AAY79983;  
 XX  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Non-typeable H. *influenzae* group 2a LB1(f) peptide N1715MEE.  
 DE  
 XX Vaccine; non-typeable *Haemophilus influenzae*; nH<sub>i</sub>; infection;  
 KW chimeric protein; *Haemophilus influenzae*; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS *Haemophilus influenzae*.  
 XX  
 PN WO9964067-A2.  
 XX  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI  
 XX WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT *Haemophilus influenzae*.  
 XX  
 XX Claim 3; Page 46; 68pp; English.  
 PS  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC *Haemophilus influenzae* strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of *Haemophilus influenzae* infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC *influenzae*. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. *influenzae* infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 22 AA;  
 SQ  
 Query Match 83.2%; Score 81.5; DB 3; Length 22;  
 Best Local Similarity 81.8%; Pred. No. 1.2e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 RSDYKLYNKNS---STLKDGLG 19  
 DB 1 RSDYKLYNKNSSSNSTLKNLGE 22  
 ||||| ||||| ||||| |||||  
 RESULT 12  
 AAY79983  
 ID AAY79983 standard; peptide; 22 AA.  
 AC AAY79983;  
 XX  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Non-typeable H. *influenzae* group 2a LB1(f) peptide N1715MEE.  
 DE  
 XX Vaccine; non-typeable *Haemophilus influenzae*; nH<sub>i</sub>; infection;  
 KW chimeric protein; *Haemophilus influenzae*; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS *Haemophilus influenzae*.  
 XX  
 PN WO9964067-A2.  
 XX  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI  
 XX WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT *Haemophilus influenzae*.  
 XX  
 XX Example 1; Page 30; 68pp; English.  
 PS  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC *Haemophilus influenzae* strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of *Haemophilus influenzae* infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC *influenzae*. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. *influenzae* infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 XX Sequence 22 AA;  
 SQ  
 Query Match 83.2%; Score 81.5; DB 3; Length 22;  
 Best Local Similarity 81.8%; Pred. No. 1.2e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 RSDYKLYNKNS---STLKDGLG 19  
 DB 1 RSDYKLYNKNSSSNSTLKNLGE 22  
 ||||| ||||| ||||| |||||  
 RESULT 13  
 AAY79974  
 ID AAY79974 standard; peptide; 22 AA.  
 AC AAY79974;  
 XX  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Non-typeable H. *influenzae* group 2 LB1(f) peptide N1715MEE.  
 DE  
 XX Vaccine; non-typeable *Haemophilus influenzae*; nH<sub>i</sub>; infection;  
 KW chimeric protein; *Haemophilus influenzae*; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS *Haemophilus influenzae*.  
 XX  
 PN WO9964067-A2.  
 XX  
 XX 16-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US011980.  
 PF  
 XX 11-JUN-1998; 98GB-00012613.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI  
 XX WPI; 2000-116457/10.  
 DR  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT *Haemophilus influenzae*.  
 XX  
 XX Example 1; Page 30; 68pp; English.  
 PS  
 XX The present invention describes antigenic P5-like fimbria subunit  
 CC



CC peptides (LBI(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX

SQ Sequence 22 AA;

Query Match 83.2%; Score 81.5; DB 3; Length 22;  
 Best Local Similarity 81.8%; Pred. No. 1.2e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNS---STLKDGLGE 19  
 |||||  
 Db 1 RSDYKLYNKNSNSTLKNLGE 22

RESULT 14

AAAB47440  
 ID AAB47440 standard; peptide; 22 AA.

XX

AC AAB47440;

XX

DT 31-OCT-2001 (first entry)

XX

DE LBI(f) containing peptide from strain nH1-1715MEE (Group 2a type).

XX

KW surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; nH1; LBI(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.

XX

OS Haemophilus influenzae.

XX

PN WO200161013-A1.

XX

PD 23-AUG-2001.

XX

PF 13-FEB-2001; 2001WO-EP001556.

XX

PR 15-FEB-2000; 2000GB-00003502.

XX

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX

DR WPI; 2001-522599/57.

XX

PT Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.

XX

PS Claim 1; Page 26; 29pp; English.

XX

CC The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these  
 CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain nH1128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection

XX Sequence 22 AA;

RESULT 16

Query Match 83.2%; Score 81.5; DB 4; Length 22;  
 Best Local Similarity 81.8%; Pred. No. 1.2e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNS---STLKDGLGE 19  
 |||||  
 Db 1 RSDYKLYNKNSNSTLKNLGE 22

RESULT 15

AAAB47444

ID AAB47444 standard; peptide; 31 AA.

XX

AC AAB47444;

XX

DT 31-OCT-2001 (first entry)

XX

DE Entire 3rd loop from strain nH1-1715MEE (Group 2a type).

XX

KW surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; nH1; LBI(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.

XX

OS Haemophilus influenzae.

XX

PN WO200161013-A1.

XX

PD 23-AUG-2001.

XX

PF 13-FEB-2001; 2001WO-EP001556.

XX

PR 15-FEB-2000; 2000GB-00003502.

XX

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX

DR WPI; 2001-522599/57.

XX

PT Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.

XX

PS Claim 2; Page 26; 29pp; English.

XX

CC The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these  
 CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain nH1128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection

SQ Sequence 31 AA;

Query Match 83.2%; Score 81.5; DB 4; Length 31;  
 Best Local Similarity 81.8%; Pred. No. 1.8e-06;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNS---STLKDGLGE 19  
 |||||  
 Db 3 RSDYKLYNKNSNSTLKNLGE 24

```

AAV79993
ID AAV79993 standard; protein; 464 AA.
AC AAV79993;
XX
XX
XX 15-MAY-2000 (first entry)
DT
DE Plasmid LPD-LB1-III protein sequence.
XX
XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
XX Haemophilus influenzae.
OS Synthetic.
XX
XX WO9964067-A2.
PN
XX
PD 16-DEC-1999.
PF
XX 28-MAY-1999; 99WO-US011980.
PR 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
PI WPI; 2000-116457/10.
XX
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
DR Haemophilus influenza.
XX
XX Example 1; Page 30; 68pp; English.
PS
XX The present invention describes antigenic P5-like fimbria subunit
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
CC Haemophilus influenzae strains. The peptides are used for diagnosis,
CC prevention, and treatment of Haemophilus influenzae infections, such as
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
CC infection. The peptides may also be used in vaccines against H.
CC influenzae. Antibodies and probes from the present invention can be used
CC for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and
CC AA291201 to AA291252, represent sequences used in the exemplification of
CC the present invention
XX
XX Sequence 464 AA;
SQ
Query Match 83.2%; Score 81.5; DB 3; Length 464;
Best Local Similarity 81.8%; Pred. No. 5.3e-05;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Oy 1 RSDYKLYNKNS---STLKDLGE 19
Db 377 RSDYKLYNKNSSTSLKDLGE 398

RESULT 17
AAV79980
ID AAV79980 standard; peptide; 20 AA.
XX
XX AAV79980;
AC
XX
XX 15-MAY-2000 (first entry)
DT
DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-495.
XX
XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
XX Haemophilus influenzae.
OS
XX
XX WO9964067-A2.
PN
XX
PD 16-DEC-1999.
PF
XX 28-MAY-1999; 99WO-US011980.
PR 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
PI WPI; 2000-116457/10.
XX
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
DR Haemophilus influenza.
XX
XX Claim 14; Fig 5; 68pp; English.
PS
XX The present invention describes antigenic P5-like fimbria subunit
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
CC Haemophilus influenzae strains. The peptides are used for diagnosis,
CC prevention, and treatment of Haemophilus influenzae infections, such as
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
CC infection. The peptides may also be used in vaccines against H.
CC influenzae. Antibodies and probes from the present invention can be used
CC for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and
CC AA291201 to AA291252, represent sequences used in the exemplification of
CC the present invention
XX
XX Sequence 464 AA;
SQ
Query Match 77.0%; Score 75.5; DB 3; Length 20;
Best Local Similarity 85.0%; Pred. No. 1.1e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Oy 1 RSDYKLYNKNS-TLKDLGE 19
Db 1 RSDYKLYNKNSDALKDLGE 20

RESULT 18
AAV79976
ID AAV79976 standard; peptide; 21 AA.
XX
XX AAV79976;
AC
XX
XX 15-MAY-2000 (first entry)
DT
DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-502.
XX
XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
XX Haemophilus influenzae.
OS
XX
XX WO9964067-A2.
PN
XX
PD 16-DEC-1999.
PF
XX 28-MAY-1999; 99WO-US011980.
PR 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.

```

XX Bakaletz LO, Cohen J, Dequeene G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenza.  
 XX Example 1; Page 30; 69pp; English.  
 XX The present invention describes antigenic P5-like fimbria subunit  
 XX peptides (LBI(f) peptides) of P5-like fimbria proteins from various  
 XX Haemophilus influenzae strains. The peptides are used for diagnosis,  
 XX prevention, and treatment of Haemophilus influenzae infections, such as  
 XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 XX infection. The peptides may also be used in vaccine against H.  
 XX influenzae. Antibodies and probes from the present invention can be used  
 XX for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 XX AA291201 to AA291252, represent sequences used in the exemplification of  
 XX the present invention

XX SQ Sequence 21 AA;  
 Query Match 76.5%; Score 75; DB 3; Length 21;  
 Best Local Similarity 81.0%; Pred. No. 1.5e-05;  
 Matches 17; Conservative 1; Mismatches 1; Indels 2; Gaps 1;  
 Qy 1 RSDYKLYNKNS--TLKDLGE 19  
 Db 1 RSDYKLYDNSSNTLKLGE 21

RESULT 19  
 AA98912  
 ID AA98912 standard; protein; 360 AA.  
 XX  
 AC AA98912;  
 XX  
 DT 11-MAY-1999 (first entry)  
 XX  
 DE RPS2 protein.  
 XX  
 KW Elicitor receptor; elicitor binding site; mould-resistant plant;  
 KW chimeric protein; signal transduction motif; disease resistance;  
 KW fungal infection.  
 XX  
 OS Unidentified.  
 XX  
 PN W09858065-A1.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-JP002719.  
 XX  
 PR 18-JUN-1997; 97JP-00161726.  
 XX  
 PA (KIRI ) KIRIN BEER KK.  
 XX  
 PI Kakitani M, Umamoto N, Tsukahara M, Ishida I;  
 XX  
 WPI; 1999-095341/08.  
 DR N-PSDB; AAX18559.  
 XX  
 PT Production of transformant plants having increased resistance to fungal  
 PT attack - comprises use of vector encoding chimeric protein containing  
 PT elicitor-binding site of elicitor receptor together with signal  
 PT transduction sequence of expression product of plant disease resistance  
 PT gene.  
 XX  
 PS Claim 6; Page 91-94; 109pp; Japanese.  
 XX  
 CC The present invention describes new DNA which encodes a chimeric protein  
 CC containing: (i) the elicitor-binding site of an elicitor receptor; and

CC (ii) the signal transduction sequence of the expression product of a gene  
 CC which can confer disease resistance to plants. The elicitor receptor is a  
 CC receptor for glucan, polygalacturonic acid, N-acetylketooligosaccharide,  
 CC elicitor, the expression product of the Avr gene of Cladosporium fulvum,  
 CC or the expression product of the nipl gene of Rhynchosporium secalis. The  
 CC signal transduction sequence may be a leucine-rich repeat, leucine  
 CC zipper, nucleic acid binding site or serine/threonine kinase domain,  
 CC preferably from the expression product of tomato Pto. Pif, Cf-2 or Cf-9  
 CC gene, rice Xa21 gene, Arabidopsis thaliana RPS2 or RPM1 gene, or flax L6  
 CC gene. Vectors from the present invention may be used for the production  
 CC of plants, including crop plants and garden plants (such as rice,  
 CC soybean, chrysanthemum, carnation or tobacco), with an increased  
 CC resistance to moulds and other fungal infections. The present sequence  
 CC represents RPS2, given in the present invention

XX SQ Sequence 360 AA;  
 Query Match 50.0%; Score 49; DB 2; Length 360;  
 Best Local Similarity 55.6%; Pred. No. 14;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 SDYKLYNKNSSTLKLGE 19  
 Db 111 ADYKLCCKVSAILKSIGE 128

RESULT 20  
 AA98915  
 ID AA98915 standard; protein; 683 AA.  
 XX  
 AC AA98915;  
 XX  
 DT 11-MAY-1999 (first entry)  
 XX  
 DE ISER protein.  
 XX  
 KW Elicitor receptor; elicitor binding site; mould-resistant plant;  
 KW chimeric protein; signal transduction motif; disease resistance;  
 KW fungal infection.  
 XX  
 OS Unidentified.  
 XX  
 PN W09858065-A1.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-JP002719.  
 XX  
 PR 18-JUN-1997; 97JP-00161726.  
 XX  
 PA (KIRI ) KIRIN BEER KK.  
 XX  
 PI Kakitani M, Umamoto N, Tsukahara M, Ishida I;  
 XX  
 WPI; 1999-095341/08.  
 DR N-PSDB; AAX18562.  
 XX  
 PT Production of transformant plants having increased resistance to fungal  
 PT attack - comprises use of vector encoding chimeric protein containing  
 PT elicitor-binding site of elicitor receptor together with signal  
 PT transduction sequence of expression product of plant disease resistance  
 PT gene.  
 XX  
 PS Claim 7; Page 68-71; 109pp; Japanese.  
 XX  
 CC The present invention describes new DNA which encodes a chimeric protein  
 CC containing: (i) the elicitor-binding site of an elicitor receptor; and  
 CC (ii) the signal transduction sequence of the expression product of a gene  
 CC which can confer disease resistance to plants. The elicitor receptor is a  
 CC receptor for glucan, polygalacturonic acid, N-acetylketooligosaccharide,  
 CC elicitor, the expression product of the Avr gene of Cladosporium fulvum,  
 CC or the expression product of the nipl gene of Rhynchosporium secalis. The  
 CC signal transduction sequence may be a leucine-rich repeat, leucine



Query Match 50.0%; Score 49; DB 2; Length 917;  
 Best Local Similarity 55.6%; Pred. No. 46;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 SDYKLYNKNSTLKDGE 19  
 :||| | : ||  
 Db 119 ADYKLCCKVSAILKSIGE 136

RESULT 23  
 AAR83440  
 ID AAR83440 standard; protein; 965 AA.  
 XX AAR83440;  
 XX  
 XX 29-MAY-1996 (first entry)  
 DT XX  
 DE A.thaliana RPS2 polypeptide from arbitrary reading frame A.  
 XX  
 XX Rps2; identification; isolation; primer; probe; plant disease;  
 KW resistance gene; recombinant; avrRpt2; transgene.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1..965  
 FT /note= "all X residues correspond to stop codons present  
 in file AAT01592"  
 FT Peptide 1..8  
 FT /label= sig\_peptide  
 FT Region 7..22  
 FT /label= hydrophobic\_region  
 FT Protein 9..909  
 FT /label= mat\_protein  
 FT Domain 31..52  
 FT /label= leucine-zipper  
 FT /note= "DNA binding motif"  
 FT Active-site 182..190  
 FT /label= kinase-1a\_motif  
 FT /note= "nucleotide binding site"  
 FT Active-site 258..262  
 FT /label= kinase-2\_motif  
 FT /note= "nucleotide binding site"  
 FT Active-site 330..335  
 FT /label= kinase-3a\_motif  
 FT /note= "nucleotide binding site"  
 FT Domain 340..356  
 FT /label= potential\_membrane-spanning\_domain  
 XX W09528423-A1.  
 XX  
 XX 26-OCT-1995.  
 XX  
 XX 13-APR-1995; 95WO-US004589.  
 XX  
 PR 13-APR-1994; 94US-00227360.  
 PR 22-SEP-1994; 94US-00310912.  
 XX  
 XX (GEO ) GEN HOSPITAL CORP.  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (USDA ) US SEC OF AGRIC.  
 PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 XX  
 XX Ausubel FM, Staskawicz BJ, Bent AF, Dahlbeck D, Katagiri F;  
 PI Kunkel BN, Mindrinos MN, Yu G, Baker B, Ellis J, Salmeron J;  
 XX  
 DR WPI; 1995-373764/48.  
 DR N-PSDB; AAT01592.  
 XX  
 XX Identification and isolation of plant disease resistance genes - using  
 FT Rps2 DNA derived oligo:nucleotide(s), also Rps2 protein, DNA and  
 FT antibody.

XX Disclosure; Fig 2; 241pp; English.  
 PS  
 CC AAR83440 is the complete amino acid sequence of the Rps2 (resistance to  
 CC Pseudomonas syringae) protein derived from reading frame A as given in  
 CC the specification. The Rps2 protein, antibody and oligonucleotide probes  
 CC and primers can be derived from the Rps2 gene sequence (AAT01592). The  
 CC oligonucleotides can be used for the identification and isolation of  
 CC plant disease resistance genes such as the tomato Prf gene which may be  
 CC used to transform a plant cell and produce a transgenic plant resistant  
 CC to bacterial pathogens carrying the avrRpt2 avirulence gene. Antibodies  
 CC may be used to screen recombinant expression libraries for Rps family  
 CC proteins  
 XX  
 SQ Sequence 965 AA;  
 Query Match 50.0%; Score 49; DB 2; Length 965;  
 Best Local Similarity 55.6%; Pred. No. 49;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 SDYKLYNKNSTLKDGE 19  
 :||| | : ||  
 Db 119 ADYKLCCKVSAILKSIGE 136

RESULT 24  
 AAW98913  
 ID AAW98913 standard; protein; 1054 AA.  
 XX  
 AC AAW98913;  
 XX  
 DT 11-MAY-1999 (first entry)  
 XX  
 DE CER protein.  
 XX  
 KW Elicitor receptor; elicitor binding site; mould-resistant plant;  
 KW chimeric protein; signal transduction motif; disease resistance;  
 KW fungal infection.  
 XX  
 OS Unidentified.  
 XX  
 PN W09858065-A1.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-JP002719.  
 XX  
 PR 18-JUN-1997; 97JP-00161726.  
 XX  
 XX (KIRI ) KIRIN BEER KK.  
 XX  
 XX Kakitani M, Umemoto N, Tsukahara M, Ishida I;  
 PI  
 XX  
 DR WPI; 1999-095341/08.  
 DR N-PSDB; AAX18560.  
 XX  
 XX Production of transformant plants having increased resistance to fungal  
 PT attack - comprises use of vector encoding chimeric protein containing  
 PT elicitor-binding site of elicitor receptor together with signal  
 PT transduction sequence of expression product of plant disease resistance  
 PT gene.  
 XX  
 XX Claim 7; Page 46-51; 109pp; Japanese.  
 PS  
 CC The present invention describes new DNA which encodes a chimeric protein  
 CC containing: (i) the elicitor-binding site of an elicitor receptor; and  
 CC (ii) the signal transduction sequence of the expression product of a gene  
 CC which can confer disease resistance to plants. The elicitor receptor is a  
 CC receptor for glucan, polygalacturonic acid, N-acetylketooligosaccharide,  
 CC elicitor, the expression product of the Avr gene of Cladosporium fulvum,  
 CC or the expression product of the nlp1 gene of Rhynchosporium secalis. The  
 CC signal transduction sequence may be a leucine-rich repeat, leucine  
 CC zipper, nucleic acid binding site or serine/threonine kinase domain,



CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 536 AA;

Query Match 49.0%; Score 48; DB 6; Length 536;  
 Best Local Similarity 71.4%; Pred. No. 35;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 YKLNKNSSTLKDL 17  
 |||||  
 Db 14 YKLNKFFDKLKL 27

RESULT 27  
 ABB61039  
 ID ABB61039 standard; protein; 200 AA.  
 XX AC ABB61039;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 9909.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL05142.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 XX interactions.

XX Disclosure; SEQ ID NO 9909; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
 CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 200 AA;

Query Match 48.0%; Score 47; DB 4; Length 200;  
 Best Local Similarity 47.4%; Pred. No. 15;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Oy 1 RSDYKLYKNKNSSTLKDLGE 19  
 |||||  
 Db 41 RSDYKLYKNKLSRREIARE 59

RESULT 28  
 ABU24397  
 ID ABU24397 standard; protein; 648 AA.

XX AC ABU24397;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by prokaryotic essential gene #9924.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Clostridium botulinum.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvekind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,  
 XX WPI; 2003-029926/02.  
 XX DR N-PSDB; ACA28267.  
 XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 XX for homologous nucleic acids required for cellular proliferation to  
 XX isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 52321; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 648 AA;

Query Match 46.9%; Score 46; DB 6; Length 648;  
 Best Local Similarity 47.4%; Pred. No. 98;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RSDYKLYNNSTLKDGE 19  
 | | | : | | | : | |  
 DB 422 RFDVLMYDNDEFFKELCE 440

RESULT 29  
 AAW29322  
 ID AAW29322 standard; protein; 1829 AA.  
 XX  
 AC AAW29322;  
 XX  
 DT 20-APR-1998 (first entry)  
 XX  
 DE DNA polymerase with 3'-5' exonuclease activity.  
 XX  
 KW TYPol Intron; DSM 10597; thermostable; DNA polymerase; 3'-5' exonuclease;  
 KW amplification.  
 XX  
 OS Thermococcus sp.

Key Location/Qualifiers  
 FH Misc-difference 1116  
 FT /label= unknown  
 FT /note= "encoded by GGN"  
 FT Misc-difference 1118  
 FT /label= unknown  
 FT /note= "encoded by NTC"  
 FT Misc-difference 1123  
 FT /label= unknown  
 FT /note= "encoded by NTG"

DE19611759-A1.  
 XX  
 PD 02-OCT-1997.

XX 25-MAR-1996; 96DE-01011759.  
 XX  
 PR 25-MAR-1996; 96DE-01011759.  
 XX  
 PA (BOE ) BOEHRINGER MANNHEIM GMBH.  
 XX  
 PI Frey B, Niehaus F, Antranikian G;

DR WPI; 1997-481494/45.  
 DR N-PSDB; AAT86433.  
 XX  
 PT Thermostable DNA polymerase from *Thermococcus* sp. TY - useful for nucleic  
 PT acid amplification.  
 XX  
 PS Claim 1; Page 5-8; 32pp; German.  
 XX  
 CC The present sequence (TYPol Intron) is a *Thermococcus* sp. TY (DSM 10597)  
 CC thermostable DNA polymerase with 3'-5' exonuclease activity. The enzyme  
 CC can specifically amplify nucleic acid fragments of up to 5 kb in high  
 CC yields, has an activity half-life of 20 minutes at 90 degrees C, has an  
 CC optimum temperature of 70-80 degrees C, has an optimum pH of 7.5,  
 CC exhibits optimum activity at a KCl concentration of 80-100 mM, is  
 CC magnesium ion-dependent and is inhibited by manganese ions  
 XX  
 SQ Sequence 1829 AA;

Query Match 46.9%; Score 46; DB 2; Length 1829;  
 Best Local Similarity 61.5%; Pred. No. 3.6e+02;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KLYNNKSTLKD 17  
 | | | | : | | |  
 DB 1161 KLYNNENPNVLKDM 1173

RESULT 30  
 ADNS9952  
 ID ADNS9952 standard; protein; 1829 AA.

AC ADNS9952;

XX 29-JUL-2004 (first entry)

DE Thermococcus strain TY DNA polymerase protein SEQ ID NO:13.  
 XX  
 KW mutant Archeal DNA polymerase; DNA polymerase; enzyme;  
 KW reverse transcriptase.

OS Thermococcus sp.

PN WO2004039947-A2.

XX 13-MAY-2004.

XX 15-AUG-2003; 2003WO-US025762.

PR 19-AUG-2002; 2002US-00223650.

PR 12-MAY-2003; 2003US-00435766.

XX (STRA-) STRATAGENE.

XX Arezi B, Hogrefe H, Sorge JA, Hansen CJ;

XX WPI; 2004-376175/35.

DR N-PSDB; ADNS9953.

XX New recombinant mutant Archeal DNA polymerase exhibiting an increased  
 FT reverse transcriptase activity, useful for reverse transcribing an RNA  
 FT template into cDNA or for amplifying an RNA template.

XX Claim 3; SEQ ID NO 13; 208pp; English.

XX The present invention describes a recombinant mutant Archeal DNA  
 CC polymerase exhibiting an increased reverse transcriptase activity, where  
 CC the wild-type form comprises an amino acid sequence selected from the 12  
 CC fully defined sequences comprising 586-1829 amino acids of SEQ ID NO:1-23  
 CC (odd numbers only). Also described: (1) a chimeric polypeptide comprising  
 CC the mutant Archeal DNA polymerase and a second polypeptide fused to the  
 CC mutant Archeal DNA polymerase; (2) an isolated polynucleotide encoding;  
 CC (a) the mutant Archeal DNA polymerase which exhibits an increased reverse  
 CC transcriptase activity, compared to a DNA polymerase encoded by a wild-



CC type polynucleotide comprising an amino acid sequence selected from SEQ  
 CC ID NO:1-23 (odd numbers only); or (b) the chimeric polypeptide; (3) a  
 CC composition comprising the mutant Archaal DNA polymerase exhibiting an  
 CC increased reverse transcriptase activity, where the wild-type form  
 CC comprises an amino acid sequence selected from SEQ ID NO:1-23 (odd  
 CC numbers only); (4) a kit comprising a mutant Archaal DNA polymerase  
 CC exhibiting an increased reverse transcriptase activity, where the wild-  
 CC type form comprises an amino acid sequence selected from SEQ ID NO:1-23  
 CC (odd numbers only), and packaging materials; (5) reverse transcribing an  
 CC RNA template; and (6) amplifying an RNA. The recombinant mutant Archaal  
 CC DNA polymerase is useful for reverse transcribing an RNA template into  
 CC cDNA. It is also useful for amplifying an RNA template. The present  
 CC sequence represents a DNA polymerase, which is used in the  
 CC exemplification of the present invention.

XX  
 SQ Sequence 1829 AA;

Query Match 46.9%; Score 46; DB 8; Length 1829;  
 Best Local Similarity 61.5%; Pred. No. 3.6e+02; Gaps 0;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0;

Oy 5 KLYNKNSSTLKD 17  
 |||||:|:|:|:  
 Db 1161 KLYNENPNVLKDM 1173

RESULT 31  
 AAO00088  
 ID AAO00088 standard; protein; 143 AA.  
 XX  
 AC AAO00088;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 13980.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US004927.  
 XX  
 PR 28-FEB-2000; 2000US-00515126.  
 PR 18-MAY-2000; 2000US-00577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 FI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-514838/56.  
 DR N-PSDB; AAI80019.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.  
 XX  
 PS Claim 20; SEQ ID NO 13980; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 143 AA;

Query Match 45.9%; Score 45; DB 4; Length 143;  
 Best Local Similarity 53.8%; Pred. No. 22;  
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Oy 4 YKLYNKNSSTLKD 16  
 :||:|:|:|:  
 Db 52 HKVISRNSQIVKD 64

RESULT 32  
 ABP30404  
 ID ABP30404 standard; protein; 311 AA.  
 XX  
 AC ABP30404;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 9984.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 FN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 PR 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; ABN71035.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX  
 PS Claim 1; Page 4127; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying



CC the European Patent Office

SQ Sequence 624 AA;

Query Match 45.9%; Score 45; DB 5; Length 624;

Best Local Similarity 50.0%; Pred. No. 1.4e+02;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSTLKDGLG 18

Db 69 RDDYEFMIANPLTWKDLG 86

RESULT 35

ADN73899  
ID ADN73899 standard; protein; 1017 AA.

XX AC ADN73899;

XX DT 15-JUN-2004 (first entry)

XX DE Thale cress protein repressed in E2Fa/Dpa expressing plants SeqID 1794.

XX KW plant; transgenic; E2Fa/Dpa transcription factor; growth regulator;

XX KW animal feed product; thale cress; cell wall biosynthesis;

XX KW nitrogen metabolism; carbon metabolism.

XX OS Arabidopsis thaliana.

XX XX WO2004035798-A2.

XX XX 29-APR-2004.

XX XX 20-OCT-2003; 2003WO-EP011658.

XX XX 18-OCT-2002; 2002EP-00079408.

XX XX (CROP-) CROPDESIGN NV.

XX XX Inze D, De Veylder L, Vlieghe K;

XX XX WPI; 2004-348466/32.

XX XX N-PSDB; ADN73898.

PT Altering plant characteristics, useful for producing plants for enzyme or pharmaceutical production comprises modifying in a plant, expression of one or more nucleic acids and/or modifying level or activity of one or more proteins.

PS Claim 1; SEQ ID NO 1794; 134pp; English.

XX This invention relates to a novel method for altering one or more plant characteristics. Specifically, it refers to identifying genes that are up - or down-regulated in transgenic plants overexpressing the heterodimeric E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to alter plant characteristics accordingly. The present invention describes generating transgenic plants for the production of growth regulators, enzymes, therapeutics, pharmaceuticals and animal feed products, where the altered plant characteristics are selected from increased yield or biomass, enhanced survival capacity, stress tolerance, plant architecture or physiology, altered endoreplication, biochemistry, signal transduction, storage lipid mobilisation and/or altered photosynthesis, each relative to the corresponding wild type plants. Accordingly, these sequences can also be useful as positive or negative selectable markers during transformation of cells or tissues. The identified genes play a role in a variety of biological processes such as DNA replication, cell wall biosynthesis, nitrogen and/or carbon metabolism or they function as transcription factors. This polypeptide sequence is thale cress protein expressed by a gene repressed 1.3 fold or more in plants overexpressing the E2Fa/Dpa transcription factor, given in an exemplification of the invention.

XX Sequence 1017 AA;

SQ

Query Match 45.9%; Score 45; DB 8; Length 1017;

Best Local Similarity 52.9%; Pred. No. 2.5e+02;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSTLKDGLG 19

Db 936 DTKLFLKSTFDIKDLGE 952

RESULT 36

ADA34312  
ID ADA34312 standard; protein; 733 AA.

XX AC ADA34312;

XX DT 20-NOV-2003 (first entry)

XX DE Acinetobacter baumannii protein #1473.

XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine; plant biocontrol agent.

XX OS Acinetobacter baumannii.

XX XX US6562958-B1.

XX XX 13-MAY-2003.

XX XX 04-JUN-1999; 99US-00328352.

XX XX 09-JUN-1998; 98US-0088701P.

XX XX (GENO-) GENOME THERAPEUTICS CORP.

XX XX Breton G, Bush D;

XX XX WPI; 2003-576092/54.

XX XX N-PSDB; ADA30186.

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

PS Example; SEQ ID NO 5599; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

SQ Sequence 733 AA;

Query Match 45.4%; Score 44.5; DB 6; Length 733;

Best Local Similarity 45.5%; Pred. No. 2.1e+02;

Matches 10; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

Qy 1 RSDYKLYNKSST-----LKDL 17

Db 1 RQDFKLMNQNTTDTIDLKEL 22

RESULT 37

ADK34379  
ID ADK34379 standard; protein; 96 AA.

XX AC ADK34379;

XX 06-MAY-2004 (first entry)  
XX Novel human polypeptide SeqID6461.  
XX antiarthritic; antiparkinsonian; neuroprotective; nootropic;  
KW immunosuppressive; cytostatic; antiparasitic; antiinflammatory;  
KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;  
KW arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;  
KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;  
KW fungus; parasite; human.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 1. 96  
XX /label= OTHER  
XX /note= "OTHER= All Xaa's in this sequence are unknown  
XX amino acids or the site of a stop codon within the DNA  
XX sequence"  
XX WO200216439-A2.  
XX  
XX 28-FEB-2002.  
XX  
XX 05-MAR-2001; 2001WO-US004941.  
XX  
XX 07-MAR-2000; 2000US-00519705.  
XX 19-MAY-2000; 2000US-00574454.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Dermanan RT;  
XX WPI; 2002-280918/32.  
XX  
XX Isolated polynucleotide encoding bone marrow derived polypeptides useful  
XX for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's  
XX disease, and inflammatory bowel disease.  
XX  
XX Claim 20; SEQ ID NO 6461; 504pp; English.  
XX  
XX This invention relates to a novel isolated polynucleotide comprising a  
XX nucleotide sequence selected from one of 1680 sequences, a mature protein  
XX coding portion of them, an active domain of them and their complementary  
XX sequences. The invention may be useful for the production of compounds  
XX with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,  
XX immunosuppressive, cytostatic, antiparasitic, antiinflammatory,  
XX antibacterial, antiviral, antifungal or antiparasitic activity. In  
XX addition, the disclosed sequences may be useful for gene therapy. The  
XX polypeptides or their antibodies are useful for treating many diseases  
XX such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,  
XX psoriasis, inflammatory bowel disease and infections caused by bacteria,  
XX viruses, fungi or parasites. The present sequence is that of a human  
XX polypeptide of the invention.  
XX  
XX Sequence 96 AA;  
XX  
XX Query Match 44.9%; Score 44; DB 5; Length 96;  
XX Best Local Similarity 64.3%; Pred. No. 20;  
XX Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
XX  
XX 4 YKLYNKNSSITLKD 17  
XX :|||:|||:|||  
XX 39 FKLYTKNYSRLHDL 52  
XX  
XX RESULT 38  
XX AAG22869  
XX ID AAG22869 standard; protein; 140 AA.  
XX  
XX AAG22869;  
XX

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 25961.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127452P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 05-MAY-1999; 99US-0132484P.  
XX 06-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 07-MAY-1999; 99US-0132487P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
XX 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.

PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160800P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 44.9%; Score 44; DB 3; Length 140;  
Best Local Similarity 53.3%; Pred. No. 32;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KLYNKNSTLKLGE 19  
:| | | | | | | | | |  
Db 60 RLTKNPNATIEDLGE 74

RESULT 39

AAM50854  
ID AAM50854 standard; protein; 550 AA.  
XX  
AC AAM50854;  
XX  
DT 29-AUG-2003 (revised)  
DT 01-MAY-2002 (first entry)  
XX  
DE Comamonas sp. cyclopentanone 1,2-monooxygenase.  
XX  
KW Cyclopentanone 1,2-monooxygenase; CPWO; Pseudomonas; enzyme.  
XX

OS Comamonas sp; 'NCIMB 9872'.

XX WO200206452-A2.

XX 24-JAN-2002.

XX 13-JUL-2001; 2001WO-CA001032.

XX 18-JUL-2000; 2000US-0218842P.

XX (CANA ) NAT RES COUNCIL CANADA.

XX Iwaki H, Hasegawa Y, Lau PCK;

XX WPI; 2002-179790/23.

XX N-PSDB; ABA91715.

XX Novel cyclopentanone monooxygenase from Comamonas having increased enzymatic activity for growing cells in a medium containing cyclopentanol or cyclopentanone as sole carbon source.

XX Claim 15; Page 27-28; 34pp; English.

XX The present sequence is that of Comamonas (formerly Pseudomonas) sp. NCIMB 9872 cyclopentanone 1,2-monooxygenase (CPMO), as deduced from an isolated genomic DNA fragment (see ABA91715). CPMO carries out the second step of a degradation pathway that allows Comamonas sp. NCIMB 9872 to use cyclopentanol as a sole C-source for growth. CPMO protein (mol.wt. 62,111) shows 36.5% homology with the cyclohexanone 1,2-monooxygenase of Acinetobacter sp. NCIMB 9871. It was expressed in Escherichia coli as an IPTG-inducible protein. Isolated CPMO DNA, recombinant expression vectors, especially plasmids and viral DNA vectors, transformed cells, especially E. coli, purified CPMO, recombinant CPMO and a method for growing cells in vitro in the presence of cyclopentanol or cyclopentanone as sole C-source, where the cells are transformed with an expression construct encoding enzymatically active CPMO, are all claimed. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 550 AA;

Query Match 44.9%; Score 44; DB 5; Length 550;

Best Local Similarity 43.8%; Pred. No. 1.8e+02;

Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YKLYNKNSSTLKDLGE 19

Db 352 YEIFNQNNVTLLVDVNE 367

RESULT 40

ABR52727

ID ABR52727 standard; protein; 1058 AA.

XX ABR52727;

XX 20-JUN-2003 (first entry)

XX Protein sequence #SEQ ID 319.

XX Multiprotein complex; eukaryote; drug target; diagnosis.

XX Saccharomyces cerevisiae.

XX EP1258494-A1.

XX 20-NOV-2002.

XX 20-DEC-2001; 2001EP-00130253.

XX 15-MAY-2001; 2001EP-00111774.

XX (CELL-) CELLZONE AG.

PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

XX Marzioch M, Schultz JD, Superti-Furga GD;

XX WPI; 2003-250078/25.

XX N-PSDB; ACC60769.

XX New isolated protein complexes useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of disease or disorder.

XX Disclosure; SEQ ID NO 319; 17pp + Sequence Listing; English.

XX The invention relates to multiprotein complexes from eukaryotes. Proteins of the invention and DNA sequences encoding them are given in records ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are obtainable by using a protein as a bait and isolating the set of proteins which is attached thereto from cells. Such protein complexes may comprise up to 30 distinct proteins. Protein complexes of the invention are useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of a disease or disorder. Note: The sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office. The complete document is available on CD-ROM

XX Sequence 1058 AA;

Query Match 44.9%; Score 44; DB 6; Length 1058;

Best Local Similarity 60.0%; Pred. No. 4e+02;

Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSTLKDL 17

Db 72 DYKAYKPNLSLNDL 86

Search completed: November 24, 2004, 09:21:45

Job time : 84.8152 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:11:44 ; Search time 18.1739 Seconds  
(without alignments)  
116.473 Million cell updates/sec

Title: US-09-719-379A-2  
Perfect score: 112  
Sequence: 1 RSDYKLYNKSSNSTLKNLGE 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: Pirl:\*  
2: Pirl2:\*  
3: Pirl3:\*  
4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 76.5  | 68.3        | 353    | 2 C64187 | outer membrane pro |
| 2          | 53    | 47.3        | 442    | 2 C84985 | trigger factor [im |
| 3          | 48.5  | 43.3        | 150    | 2 T28904 | hypothetical prote |
| 4          | 48    | 42.9        | 706    | 2 A80022 | probable membrane  |
| 5          | 47    | 42.0        | 200    | 2 S15111 | hemorrhagic factor |
| 6          | 47    | 42.0        | 308    | 2 A70254 | hypothetical prote |
| 7          | 47    | 42.0        | 1374   | 2 S69875 | hypothetical prote |
| 8          | 46    | 41.1        | 324    | 2 T02109 | calmodulin-like pr |
| 9          | 46    | 41.1        | 332    | 2 B71606 | probable integral  |
| 10         | 46    | 41.1        | 700    | 2 E86397 | protein T7N9_14 [i |
| 11         | 46    | 41.1        | 2052   | 2 C97038 | phage-related prot |
| 12         | 46    | 41.1        | 4767   | 2 T31345 | hypothetical prote |
| 13         | 45.5  | 40.6        | 103    | 2 D24444 | hypothetical prote |
| 14         | 45    | 40.2        | 182    | 2 T24672 | hypothetical prote |
| 15         | 45    | 40.2        | 234    | 2 S36348 | opacity protein op |
| 16         | 45    | 40.2        | 238    | 2 A46130 | fiber protein E6 ( |
| 17         | 45    | 40.2        | 241    | 2 S65082 | fiber protein E6 ( |
| 18         | 45    | 40.2        | 246    | 2 S65063 | hypothetical prote |
| 19         | 45    | 40.2        | 285    | 2 C97178 | hypothetical prote |
| 20         | 45    | 40.2        | 338    | 2 S45908 | hypothetical prote |
| 21         | 45    | 40.2        | 430    | 2 T28318 | ORF MSV157 hypote  |
| 22         | 45    | 40.2        | 481    | 2 T18455 | hypothetical prote |
| 23         | 45    | 40.2        | 589    | 2 A71918 | probable secretion |
| 24         | 45    | 40.2        | 611    | 2 T27013 | hypothetical prote |
| 25         | 45    | 40.2        | 1382   | 2 S70310 | hypothetical prote |
| 26         | 45    | 40.2        | 1619   | 2 T18499 | hypothetical prote |
| 27         | 45    | 40.2        | 1681   | 2 S59693 | hypothetical prote |
| 28         | 45    | 40.2        | 1758   | 2 S57015 | probable purine nu |
| 29         | 45    | 40.2        | 1796   | 2 S65004 | probable membrane  |

|    |      |      |      |          |                    |
|----|------|------|------|----------|--------------------|
| 30 | 45   | 40.2 | 1859 | 2 S63325 | probable membrane  |
| 31 | 45   | 40.2 | 1859 | 2 S64633 | probable membrane  |
| 32 | 44.5 | 39.7 | 1237 | 2 A54080 | protein-tyrosine-p |
| 33 | 44   | 39.3 | 191  | 2 C75299 | tellurium resistan |
| 34 | 44   | 39.3 | 252  | 2 H97060 | uncharacterized co |
| 35 | 44   | 39.3 | 262  | 2 T24854 | hypothetical prote |
| 36 | 44   | 39.3 | 312  | 2 T32211 | hypothetical prote |
| 37 | 44   | 39.3 | 327  | 2 T23572 | hypothetical prote |
| 38 | 44   | 39.3 | 359  | 2 A90525 | peptide chain rele |
| 39 | 44   | 39.3 | 451  | 2 A23535 | clustered asparagi |
| 40 | 44   | 39.3 | 477  | 2 T52382 | zinc finger protei |
| 41 | 44   | 39.3 | 528  | 2 G90569 | hypothetical prote |
| 42 | 44   | 39.3 | 529  | 2 S46116 | probable regulator |
| 43 | 44   | 39.3 | 739  | 2 E97245 | beta-lactamase cla |
| 44 | 44   | 39.3 | 2269 | 2 T18472 | hypothetical prote |
| 45 | 43.5 | 38.8 | 173  | 2 G86824 | hypothetical prote |

ALIGNMENTS

RESULT 1

C64187  
outer membrane protein A homolog - Haemophilus influenzae (strain Rd KW20)  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C;Accession: C64187  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: C64187  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-353 <TIGR>  
A;Cross-references: UNIPROT:P43840; GB:U32796; GB:L42023; NID:G1574086; PIDN:AAC22819.1;  
C;Superfamily: outer membrane protein A

Query Match 68.3%; Score 76.5; DB 2; Length 353;  
Best Local Similarity 77.3%; Pred. No. 0.0017;  
Matches 17; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22  
|||||||:|  
Db 132 RSDYKLYNENS---STLKLGE 150

RESULT 2

C84985  
trigger factor [imported] - Buchnera sp. (strain APS)  
C;Species: Buchnera sp.  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 12-Jun-2003  
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
Nature 407, 81-86, 2000  
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A  
A;Reference number: A84930; MUID:20445173; PMID:10593077  
A;Accession: C84985  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-442 <STO>  
A;Cross-references: GB:AP000398; GSPDB:GN00144  
A;Experimental source: strain APS  
C;Genetics:  
A;Gene: tig; BU474  
C;Superfamily: FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor); BKBP-type

Query Match 47.3%; Score 53; DB 2; Length 442;  
Best Local Similarity 56.2%; Pred. No. 5.4;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 5 KLYNKNSSNSTLKNL 20  
Db 388 KLYNKNKLNKNTMKVI 403

RESULT 3  
T28904  
hypothetical protein T25B6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28904  
R:Favellio, A.  
submitted to the EMBL Data Library, December 1995  
A:Description: The sequence of C. elegans cosmid T25B6.  
A:Reference number: Z20541  
A:Accession: T28904  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-150 <FAV>  
A:Cross-references: UNIPROT:Q22762; EMBL:U41546; PIDN:AAC48218.1; GSPDB:GN000028; CESP:T2  
A:Experimental source: strain Bristol N2; clone T25B6  
C:Genetics:  
A:Gene: CESP:T25B6.3  
A:Map position: X  
A:Introns: 25/3; 85/2

Query Match 43.3%; Score 48.5; DB 2; Length 150;  
Best Local Similarity 46.2%; Pred. No. 8;  
Matches 12; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

Qy 1 RSDYKLYNK-----NSSNSTLKN 19  
Db 90 QSDHLYNKYFFHLGHNFLSNTCKN 115

RESULT 4  
AB0022  
probable membrane protein YP00174 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AB0022  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AB0022  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-706 <KUR>  
A:Cross-references: UNIPROT:Q8ZJD9; GB:AL590842; PIDN:CAC89036.1; PID:g15978277; GSPDB:G  
C:Genetics:  
A:Gene: YP00174

Query Match 42.9%; Score 48; DB 2; Length 706;  
Best Local Similarity 40.0%; Pred. No. 47;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSNSTLKNLGE 22  
Db 543 DYERFNVKASNTVLSLNLQ 562

RESULT 5  
S15111  
hemorrhagic factor LHFII (EC 3.4.24.-) - bushmaster  
C:Species: Lachesis muta muta (bushmaster)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C:Accession: S15111  
R:Sanchez, E.F.; Diniz, C.R.; Richardson, M.  
FEBS Lett. 282, 178-182, 1991

A:Title: The complete amino acid sequence of the haemorrhagic factor LHFII, a metallopro  
A:Reference number: S15111; MUID:91224327; PMID:2026257  
A:Accession: S15111  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-200 <FEB>  
A:Cross-references: UNIPROT:P22796  
C:Superfamily: atrolysin C  
C:Keywords: hydrolase; metalloproteinase; venom; zinc  
F:140,144,150/Binding site: zinc (His) #status predicted  
F:141/Active site: Glu #status predicted

Query Match 42.0%; Score 47; DB 2; Length 200;  
Best Local Similarity 47.6%; Pred. No. 18;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 58 SNOQLINVQSAANDTLKTFGE 78

RESULT 6  
A70254  
hypothetical protein BBK23 - Lyme disease spirochete plasmid K/lp36  
C:Species: Borrelia burgdorferi (lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: A70254  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: A70254  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-308 <KLE>  
A:Cross-references: UNIPROT:O50827; GB:AE000788; NID:g2690123; PIDN:AAC66174.1; PID:g269  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid

Query Match 42.0%; Score 47; DB 2; Length 308;  
Best Local Similarity 60.0%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 6 LYNKNSSNSTLKNL 20  
Db 183 IYNKNSTIENSFLKKI 197

RESULT 7  
S69875  
hypothetical protein YML133c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein L0519; hypothetical protein L0532; hypothetical  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: S69875; S58194; S58195; S50953; S50954; S64818  
R:Bowman, S.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: S58194  
A:Accession: S69875  
A:Molecule type: DNA  
A:Residues: 1-1374 <BOW>  
A:Cross-references: UNIPROT:Q03099; EMBL:Z50178; MIPS:YML133c  
A:Experimental source: strain AB972  
A:Accession: S58194  
A:Molecule type: DNA  
A:Residues: 'PRL', 248, 'K', 250, 'IDELTVIRLP', 261, 'SR', 266-1374 <BOF>  
A:Cross-references: EMBL:Z50178; NID:g27528; PIDN:CAA90549.1; PID:s190749; PID:g1326004  
A:Experimental source: strain AB972  
A:Note: the difference at the amino end is due to translation from the presumed intron s



A:Accession: S58195  
A:Molecule type: DNA  
A:Residues: 1-265, HVCCPS: <BOA>  
A:Cross-references: EMBL:Z50178; NID:G927528; PIDN:CAA90550.1; PID:G927529  
A:Experimental source: strain AB972  
A:Note: the difference at the carboxyl end is due to translation from the presumed intron  
R:Wedler, H.; Wambutt, R.  
submitted to the EMBL Data Library, January 1995  
A:Description: Sequence of a 37 kb DNA fragment from chromosome XII of *Saccharomyces cerevisiae*  
A:Reference number: S50950  
A:Accession: S50953  
A:Molecule type: DNA  
A:Residues: 170-278, 'G', 280-329, 'D', 331-432, 'G', 434-824, 'N', 826-1072, 'G', 1074-1374 <WED>  
A:Cross-references: EMBL:Z47973; NID:G642313; PIDN:CAA87990.1; PID:G642317  
A:Accession: S50954  
A:Molecule type: DNA  
A:Residues: 1-265, 'HV', 766-767, 'P', 769 <WED>  
A:Cross-references: EMBL:Z47973; NID:G642313; PIDN:CAA87989.1; PID:G642316  
R:Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64792  
A:Accession: S64818  
A:Molecule type: DNA  
A:Residues: 170-278, 'G', 280-329, 'D', 331-432, 'G', 434-824, 'N', 826-1072, 'G', 1074-1374 <WED>  
A:Cross-references: EMBL:Z71171; NID:G1360281; PIDN:CAA97520.1; PID:G136028  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 13L; 12L  
A:Introns: 265/2  
A:Note: Closely related hypothetical proteins containing similar 12-residue repeats are  
C:Keywords: tandem repeat; transmembrane protein  
F:414-430/Domain: transmembrane #status predicted <TMM>  
F:847-966/Region: 12-residue repeats (T-T-T-[EKA]-[SG]-[TI]-[ND]-[ASV]-[TRSNK]-T-[NS]-[A

Query Match 42.0%; Score 47; DB 2; Length 1374;  
Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 YNKNSSNSTLKNL 20  
|||||  
Db 275 YNKNSSSEPRCLKL 288

RESULT 8  
T02109  
calmodulin-like protein At2g41100 [imported] - Arabidopsis thaliana  
N:Alternate names: calmodulin-related protein T3K9.13  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T02109; G84837  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, M.; Koo, H.; Moffat, K.S.; Cronin, L.S.; Shen, M.; Vanaken, S.B.; Unayam, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Reference number: 214570  
A:Accession: T02109  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-324 <ROU>  
A:Cross-references: UNIPROT:P25071; EMBL:AC004261; NID:G3402695; PID:G3402707  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.S.; Shen, M.; Vanaken, S.B.; Unayam, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: AB4420; MUID:20083487; PMID:10617137  
A:Accession: G84837  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <STO>  
A:Cross-references: GB:AE002093; NID:G3402707; PIDN:AAD12001.1; GSPDB:GN00139  
C:Genetic9:  
A:Gene: TCH3; T3K9.13; At2g41100

A:Map position: 2  
A:Introns: 26/1; 115/1; 205/1

Query Match 41.1%; Score 46; DB 2; Length 324;  
Best Local Similarity 38.1%; Pred. No. 41;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLG 21  
Db 14 RESFLFDKNGDGSITKKELG 34

RESULT 9  
B71606  
probable integral membrane protein PFB0790c - malaria parasite (*Plasmodium falciparum*)  
C:Species: *Plasmodium falciparum*  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: B71606  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; ; Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71606  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-332 <GAR>  
A:Cross-references: UNIPROT:O96251; GB:AE001418; GB:AE001362; NID:G3845275; PIDN:AAC7194  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0790C

Query Match 41.1%; Score 46; DB 2; Length 332;  
Best Local Similarity 57.1%; Pred. No. 42;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 7 YNKNSSNSTLKNL 20  
|||||  
Db 265 YNNSSNNNTSNNI 278

RESULT 10  
E86397  
protein T7N9.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: E86397  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86397  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-700 <STO>  
A:Cross-references: UNIPROT:Q9LFX9; GB:AE005172; NID:G8778857; PIDN:AAF79856.1; GSPDB:GN0139  
C:Genetics:  
A:Gene: T7N9.14  
A:Map position: 1

Query Match 41.1%; Score 46; DB 2; Length 700;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLG 21  
|||||

Db 139 SDFKPIFTVSSNETLEKLG 158

RESULT 11

C97038

phage-related protein, Yqbo B. subtilis homolog [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004

C;Accession: C97038

R;Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4836, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: C97038

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2052 <KUR>

A;Cross-references: UNIPROT:Q97K01; GB:AE001437; PIDN:AAK79094.1; PID:gl5024039; GSPDB:C97038

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC1120

Query Match 41.1%; Score 46; DB 2; Length 2052;

Best Local Similarity 55.6%; Pred. No. 2.7e+02;

Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 7 YNKSS--SNSTLNKLG 22

Db 1646 FEKSSNFKNLKNIGD 1663

RESULT 12

T31345

hypothetical protein G01D9.5 - Caenorhabditis briggsae

C;Species: Caenorhabditis briggsae

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

A;Reference number: T31345

R;Waterston, R.

submitted to the EMBL Data Library, April 1996

A;Description: The C. briggsae genome sequencing project.

A;Reference number: 221010

A;Accession: T31345

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-4767 <WAT>

A;Cross-references: UNIPROT:Q17301; EMBL:U56248; NID:gl293789; PID:gl293790; PIDN:AAA986

C;Genetics:

A;Introns: 269/3; 341/3; 853/1; 920/2; 4452/3; 4534/3; 4592/3; 4654/3; 4670/3; 4707/2; 4707/2; 4707/2

A;Note: G01D9.5

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;3472-3541/Domain: acyl carrier protein homology <ACPI>

F;4039-4427/Domain: acetate-CoA ligase homology <ACLI>

F;4447-4514/Domain: acyl carrier protein homology <ACP2>

F;2210,3505/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 41.1%; Score 46; DB 2; Length 4767;

Best Local Similarity 47.4%; Pred. No. 6.5e+02;

Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSTLNKL 20

Db 3115 SDFQYDKNIGSTQTSKIL 3133

RESULT 13

D24444

hypothetical protein 4 (16S-23S rRNA spacer region) - Chlorella ellipsoidea chloroplast

C;Species: chloroplast Chlorella ellipsoidea

C;Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 09-Jul-2004

C;Accession: D24444

R;Yamada, T.; Shimaji, M.

Nucleic Acids Res. 14, 3827-3839, 1986

A;Title: Peculiar feature of the organization of rRNA genes of the Chlorella chloroplast

A;Reference number: A93622; MUID:86232622; PMID:3714498

A;Accession: D24444

A;Molecule type: DNA

A;Residues: 1-103 <YAM>

A;Cross-references: UNIPROT:P05721

C;Genetics:

A;Genome: chloroplast

C;Superfamily: Chlorella ellipsoidea chloroplast hypothetical protein 4 (16S-23S rRNA sp

C;Keywords: chloroplast

Query Match 40.6%; Score 45.5; DB 2; Length 103;

Best Local Similarity 48.0%; Pred. No. 15;

Matches 12; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

Qy 3 DY-----KLYNKSSNSTLNKLG 22

Db 29 DYKLIKKLSKLTNSTKANLGD 53

RESULT 14

T24672

hypothetical protein T07G12.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T24672

R;Cummings, P.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19921

A;Accession: T24672

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-182 <WIL>

A;Cross-references: UNIPROT:O18058; EMBL:Z82282; PIDN:CAB05271.1; GSPDB:GN00022; CESP:T.O

A;Experimental source: clone T07G12

C;Genetics:

A;Gene: CESP:T07G12.1

A;Map position: 4

A;Introns: 102/2; 147/3

C;Superfamily: calmodulin; calmodulin repeat homology

C;Keywords: EF hand

Query Match 40.2%; Score 45; DB 2; Length 182;

Best Local Similarity 36.8%; Pred. No. 31;

Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKLYNKSSNSTLNKLG 22

Db 29 FKLFDKGNNTMNIKELGE 47

RESULT 15

S36348

opacity protein opa65 - Neisseria gonorrhoeae (strain VP1) (fragment)

N;Alternate names: outer membrane protein opa65

C;Species: Neisseria gonorrhoeae

A;Variety: strain VP1

C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

C;Accession: S36348; S28617

R;Kupsch, E.M.; Knepper, B.; Kuroki, T.; Heuer, I.; Meyer, T.F.

EMBO J. 12, 641-650, 1993

A;Title: Variable opacity (Opa) outer membrane proteins account for the cell tropisms di

A;Reference number: S36328; MUID:93178439; PMID:8440254

A;Accession: S36348

A;Molecule type: DNA

A;Residues: 1-234 <KUP>

A;Cross-references: UNIPROT:Q04885; EMBL:Z18940; NID:g49336; PIDN:CAA79373.1; PID:g94080

A;Experimental source: strain VP1

A;Note: expression of opacity proteins is regulated by the number of translated repeat

of repeats place the start codon in frame with the rest of the protein

C;Genetics:

A;Gene: opa65

C;Superfamily: opacity protein

100

Best Local Similarity 90.0%; Pred. No. 50;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KLYNKNSSN 14  
| | | | | | | |  
Db 122 KLYNTNSSN 131

RESULT 20  
S45908  
hypothetical protein YBR050c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YBR0504  
C:Species: Saccharomyces cerevisiae  
C>Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 09-Jul-2004  
C:Accession: S45908; S45905; S55849  
R:Aljinovic, G.; Pohl, F.M.; Pohl, T.M.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45906  
A:Accession: S45908  
A:Molecule type: DNA  
A:Residues: 1-338 <ALJ>  
A:Cross-references: UNIPROT:P38232; EMBL:Z35919; NID:G536281; PIDN:CAA84994.1; PID:G5362  
A:Experimental source: strain S288C  
R:Aljinovic, G.  
submitted to the EMBL Data Library, October 1994  
A:Description: Sequence and analysis of 24 kb on chromosome II of Saccharomyces cerevisiae  
A:Reference number: S49503  
A:Accession: S49505  
A:Molecule type: DNA  
A:Residues: 1-338 <AL2>  
A:Cross-references: EMBL:Z46260; NID:G559942; PIDN:CAA86393.1; PID:G559946  
A:Experimental source: strain S288C  
R:Aljinovic, G.; Pohl, T.M.  
Yeast 11, 475-479, 1995  
A:Title: Sequence and analysis of 24 kb on chromosome II of Saccharomyces cerevisiae.  
A:Reference number: S55846; MUID:95321020; PMID:7597852  
A:Accession: S55849  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-338 <ALW>  
A:Cross-references: EMBL:Z46260; NID:G559942; PIDN:CAA86393.1; PID:G559946  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994  
C:Genetics:  
A:Gene: SGD:REG2  
A:Cross-references: SGD:S0000254; MIPS:YBR050c  
A:Map position: 2R  
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YBR050c

Query Match 40.2%; Score 45; DB 2; Length 338;  
Best Local Similarity 58.8%; Pred. No. 59;  
Matches 10; Conservative 3; Mismatches 0; Indels 4; Gaps 1;

Qy 7 YNKNKNS----SNNSTLKN 19  
| | | | | | | |  
Db 109 YNKNKHKVRSSNNTVKN 125

RESULT 21  
T28318  
ORF MSV157 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28318  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28318  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-430 <AFO>  
A:Cross-references: UNIPROT:Q9YVT5; EMBL:AF063866; NID:G4049647; PIDN:AAC97678.1; PID:G4  
C:Genetics:

A:Note: MSV157

Query Match 40.2%; Score 45; DB 2; Length 430;  
Best Local Similarity 88.9%; Pred. No. 76;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DYKLYNKNKNS 11  
: | | | | | | | |  
Db 321 NYKLYNKNKNS 329

RESULT 22  
T18465  
hypothetical protein CO475c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18465  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18465  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-481 <LAW>  
A:Cross-references: UNIPROT:O77353; EMBL:AL008970; NID:e1407852; PIDN:CAA15608.1  
C:Genetics:  
A:Map position: 3  
A:Note: CO475c

Query Match 40.2%; Score 45; DB 2; Length 481;  
Best Local Similarity 41.2%; Pred. No. 86;  
Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Qy 6 LYNNKSSNSTLKNLGE 22  
: | | | | | | | |  
Db 86 IYNNKSSNNLTKDVDD 102

RESULT 23  
A71918  
Probable secretion/efflux ABC transporter, ATP-binding protein - Helicobacter pylori (st  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: A71918  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: A71918  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-589 <ARN>  
A:Cross-references: UNIPROT:Q9ZLN1; GB:AE001487; GB:AE001439; NID:G4155086; PIDN:AAD0613  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0547  
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
F;363-562/Domain: ATP-binding cassette homology <ABC>

Query Match 40.2%; Score 45; DB 2; Length 589;  
Best Local Similarity 58.8%; Pred. No. 1.1e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 4 YKLYNKNSSNSTLKNL 20  
| | | | | | | |  
Db 376 YALVGNASGKSTLINL 392

RESULT 24  
T27013  
hypothetical protein Y48E1B.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T27013  
R;McMurray, A.  
submitted to the EMBL Data Library, March 1997  
A;Reference number: Z20299  
A;Accession: T27013  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A;Residues: 1-611 <WIL>  
A;Cross-references: UNIPROT:O18198; EMBL:Z93393; PIDN:CAB07690.1; GSPDB:GN000020; CESP:Y4  
A:Experimental source: clone Y48E1B  
C;Genetics:  
A;Gene: CESP:Y48E1B.4  
A;Map position: 2  
A;Introns: 59/2; 80/3; 144/3; 413/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein Y48E1B.4

Query Match 40.2%; Score 45; DB 2; Length 611;  
Best Local Similarity 40.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYLKYNKSSNSTLKNLG 21  
          ::|::|::|::|::|::|::|  
Db 97 NDLYNRLTLATRFENSG 116

RESULT 25  
S70310  
hypothetical protein YLR466w - yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae  
C;Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: S70310  
R;Wedler, H.; Wambutt, R.  
submitted to the Protein Sequence Database, August 1995  
A;Reference number: S70306  
A;Accession: S70310  
A:Molecule type: DNA  
A;Residues: 1-1382 <WED>  
A;Cross-references: UNIPROT:O13559; MIPS:YLR466w  
A;Note: experimental\_source strain S288C  
C;Genetics:  
A;Gene: SGD:YRF1-4  
A;Cross-references: SGD:S0004458  
A;Map position: 12R  
A;Note: Closely related hypothetical proteins containing similar 12-residue repeats are  
C;Keywords: tandem repeat  
F;835-974/Region: 12-residue repeats (T-T-T-[EKA] - [SG] - [TI] - [ND] - [ASV] - [TRSNK] - T - [NS] - [A]

Query Match 40.2%; Score 45; DB 2; Length 1382;  
Best Local Similarity 58.8%; Pred. No. 2.5e+02;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YKLYNKSSNSTLKNL 20  
          | | | | | | | | | |  
Db 125 YTDYKNKSSEPRKLKT 141

RESULT 26  
T18499  
hypothetical protein C0770c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18499  
R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z18935  
A;Accession: T18499  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A;Residues: 1-1619 <LAW>  
A;Cross-references: UNIPROT:O77382; EMBL:Z98551; NID:e1331903; PID:e1331920; PIDN:CAB111  
C;Genetics:

A;Residues: 'MYQYVENFNFREYVGRGFYFACKAQFDNLWGELNLCFYNPTVVDIANLILNRNREVLFRPKRGIDE', 'YLNDSR  
A;Cross-references: EMBL:Z48148; NID:G663231; PID:G663232  
A;Note: the differences at the amino end are due to translation of intron sequence  
R;Vandenbol, M.; Durand, P.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.  
Yeast 10, 1657-1662, 1994  
A;Title: Sequence analysis of a 40.2 kb DNA fragment located near the left telomere of y  
A;Reference number: S50701; MUID:95242842; PMID:7725802  
A;Accession: S50702  
A;Molecule type: DNA  
A;Residues: 'MYQYVENFNFREYVGRGFYFACKAQFDNLWGELNLCFYNPTVVDIANLILNRNREVLFRPKRGIDS', 'YLNDSR  
A;Cross-references: EMBL:Z34098  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994  
A;Note: neither the complete nucleic acid sequence nor the complete translation are show  
R;Vandenbol, M.; Durand, P.; Dion, C.; Bolle, P.; Portetelle, D.; Hilger, F.  
submitted to the EMBL Data Library, June 1994  
A;Description: Sequence analysis of a 40.1 kb DNA fragment located near the left telomere  
A;Reference number: S45146  
A;Accession: S45147  
A;Molecule type: DNA  
A;Residues: 'MYQYVENFNFREYVGRGFYFACKAQFDNLWGELNLCFYNPTVVDIANLILNRNREVLFRPKRGIDE', 'YLNDSR  
A;Cross-references: EMBL:Z34098; NID:G496934; PID:G496936  
A;Note: the differences at the amino end are due to translation of intron sequence  
R;Iye, G.; Bowman, S.; Churcher, C.  
submitted to the EMBL Data Library, December 1994  
A;Reference number: S50349  
A;Accession: S50349  
A;Molecule type: DNA  
A;Residues: 1-727,'C',729-1758 <LYE>  
A;Cross-references: GB:Z47047; EMBL:Z46921; NID:G603997; PID:G603998; MIPS:YIIL177C  
R;Louis, E.J.; Haber, J.E.  
Genetics 1331, 559-574, 1992  
A;Title: The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevis  
A;Reference number: S31214  
A;Accession: S31214  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-73,'I',75-84,'Q',86-348,'T',350-387,'VC',944-946,'FNR',950,'Y',952,'RGFYFA  
SYSLPLVLSATY',955,'SMIQQ',961,'A',963,'NCT',967-968,'T',970,'RT',973,'HRSL',978,'LRR',98  
A;Cross-references: GB:U23472; EMBL:MS8718; NID:G775202; PIDN:AAC48994.1; PID:G775205  
R;Pryde, F.E.; Huckle, T.C.; Louis, E.J.  
Yeast 11, 371-382, 1995  
A;Title: Sequence analysis of the right end of chromosome XV in Saccharomyces cerevisiae  
A;Reference number: S59000; MUID:95304851; PMID:7785338  
A;Accession: S59002  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-73,'I',75-84,'Q',86-348,'T',350-387,'VC',944-946,'FNR',950,'Y',952,'RGFYFA  
SYSLPLVLSATY',955,'SMIQQ',961,'A',963,'NCT',967-968,'T',970,'RT',973,'HRSL',978,'LRR',98  
A;Cross-references: EMBL:U23472; NID:G775202; PIDN:AAC48994.1; PID:G775205  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995  
R;Favella, T.  
submitted to the EMBL Data Library, June 1994  
A;Description: The sequence of S. cerevisiae cosmid 9196.  
A;Reference number: S46794  
A;Accession: S48919  
A;Molecule type: DNA  
A;Residues: 1-73,'I',75,'V',77-84,'Q',86-200,'H',202-241,'T',243-263,'V',265-266,'VCCPS'  
A;Cross-references: EMBL:U1583; NID:G2289854; PID:G2289878; GSPDB:GN00008; MIPS:YHL049C  
R;Miller, N.  
submitted to the EMBL Data Library, April 1995  
A;Description: The sequence of S. cerevisiae cosmid 9677.  
A;Reference number: S58816  
A;Accession: S69476  
A;Molecule type: DNA  
A;Residues: 'MEIENEICTCAIQLHLNLSLIIFSDDKTETGQSFYIDGFLVKHNNQHTIVNFTYKNK',1-73,'I',  
A;Cross-references: EMBL:U25841; EMBL:Z73537; NID:G1370583; PID:G246990; PID:G1370584  
A;Note: the nucleotide sequence was translated assuming an intron after the first base c  
A;Accession: S58831  
A;Molecule type: DNA  
A;Residues: 1-73,'I',75-84,'Q',86-150,'PCPG',439,'LS',442,'NSECQ',448,'ERQL',453-455,'YF  
A;Cross-references: EMBL:U25841; NID:G786295; PID:G786297  
R;Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.

submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64792  
A;Accession: S65339  
A;Molecule type: DNA  
A;Residues: 'ICTCIAQLHLNLSLIIFSDDKTETGQSFYIDGFLVKHNNQHTIVNFTYKNK',1-73,'I',75-84,'  
A;Cross-references: EMBL:Z73537; MIPS:YPR202W  
A;Experimental source: strain S288C (AB972)  
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanu  
submitted to the EMBL Data Library, May 1995  
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce  
A;Reference number: S56186  
A;Accession: S56191  
A;Molecule type: DNA  
A;Residues: 1-65,'N',67-73,'I',75-84,'Q',86-150,'PCPG',439,'LS',442,'NSECQ',448,'ERQL',4  
A;Cross-references: EMBL:DS0617; NID:G836685; PIDN:BAA09177.1; PID:G1009815; PID:G836691  
R;Murakami, Y.  
submitted to the EMBL Data Library, December 1994  
A;Reference number: S62230  
A;Accession: S62273  
A;Molecule type: DNA  
A;Residues: 1-65,'N',67-73,'I',75-84,'Q',86-150,'PCPG',439,'LS',442,'NSECQ',448,'ERQL',4  
A;Cross-references: EMBL:D44603; NID:G871957; PIDN:BAA08055.1; PID:G1008646; PID:G871962  
R;Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A;Description: The sequence of S. cerevisiae cosmids 9669, 9163, 9132, 8334, 8199, and 1  
A;Reference number: S50514  
A;Accession: S50514  
A;Molecule type: DNA  
A;Residues: 1-104,'NK',912,'RDGLH',918-919,'S',921,'AYF',925,'IQ' <DIE>  
A;Cross-references: EMBL:U18795; NID:G603241; PID:G603243; GSPDB:GN00005; MIPS:YEL075C  
A;Accession: S50692  
A;Molecule type: DNA  
A;Residues: 1-73,'I',75-84,'Q',86-104,'NK',912,'RDGLH',918-919,'S',921,'AYF',925,'IQ' <D  
A;Cross-references: EMBL:U18922; NID:G603405; PID:G603430; GSPDB:GN00005; MIPS:YER189W  
R;Kirsten, J.  
submitted to the EMBL Data Library, March 1995  
A;Description: The sequence of S. cerevisiae cosmid 9122.  
A;Reference number: S59414  
A;Accession: S69325  
A;Molecule type: DNA  
A;Residues: 1-64 <KIR>  
A;Cross-references: EMBL:U22383; GSPDB:GN00012; MIPS:YLR462W  
C;Genetics:  
A;Gene: MIPS:YHL049C; MIPS:YEL075C; MIPS:YER189W; MIPS:YLR462W  
A;Cross-references: SGD:S0003760  
A;Map position: 5L; 5R; 6L; 8L; 9L; 10L; 12R; 15R; 16R  
A;Introns: 387/3  
A;Note: Closely related hypothetical proteins containing similar 12-residue repeats are  
C;Keywords: nucleotide binding; P-loop; tandem repeat; transmembrane protein  
F;681-688/Region: nucleotide-binding motif A (P-loop)  
F;707-723/Domain: transmembrane #status predicted <TM>  
F;1184-1339/Region: 12-residue repeat (T-T-T-[EKA]-[SG]-[TI]-[ND]-[ASV]-[TRSNA]-T-[NS]-

Query Match 40.2%; Score 45; DB 2; Length 1758;  
Best Local Similarity 58.8%; Pred. No. 3.3e+02;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 YKLYNKNSSNSTLKNL 20  
Db 410 YTDYNNKSGSEPRUKTL 426

RESULT 29  
S65004  
Probable membrane protein YLR467W - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein D3703.4; hypothetical protein I9040; hypothetical  
C;Species: Saccharomyces cerevisiae  
C;Date: 01-Aug-1995 #sequence revision 24-May-1996 #text\_change 09-Jul-2004  
C;Accession: S65004; S67814; S62024  
R;Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64792  
A;Accession: S65004

A;Molecule type: DNA  
A;Residues: 1-1796 <WED>  
A;Cross-references: UNIPROT:P24088; EMBL:Z73327; MIPS:YLR467w; NID:g1360589; PID:e245588  
A;Experimental source: strain S288C  
A;Genetics: CH12  
R;Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S67608  
A;Accession: S67814  
A;Molecule type: DNA  
A;Residues: 1-1796 <WAM>  
A;Cross-references: EMBL:Z74389; MIPS:YDR545w; NID:g1431573; PID:e253351; PID:g1431574  
A;Experimental source: strain S288C  
A;Genetics: CH4  
R;Dietrich, F.S.; Mulligan, J.; Allen, E.; Araujo, R.; Aviles, E.; Berno, A.; Carpenter, H.; Lin, D.; Mosedale, D.; Nakahara, K.; Namath, A.; Oefner, P.; Oh, C.; Petel, F.X.; Rost, W.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: S62017  
A;Accession: S62024  
A;Molecule type: DNA  
A;Residues: 1-228 <DIE>  
A;Cross-references: EMBL:U43834; NID:g1165292; PID:g1165300  
A;Experimental source: strain AB972  
A;Genetics: CH4  
C;Genetics: <CH12>  
A;Map position: 12R  
A;Note: YLR467w  
C;Genetics: <CH4>  
A;Map position: 4R  
A;Note: YDR545w  
C;Keywords: tandem repeat; transmembrane protein  
F;836-852/Domain: transmembrane #status predicted <TMM>  
F;1269-1388/Region: 12-residue repeats (T-T-T-[EKA]) - [SG] - [TI] - [ND] - [ASV] - [TRSNNK] - T- [NS] -  
Query Match 40.2%; Score 45; DB 2; Length 1796;  
Best Local Similarity 58.8%; Pred. No. 3.3e+02;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 4 YKLYNNKSSNSTLKNL 20  
Db 539 YTDYNNKSSSPRLKTL 555  
RESULT 30  
S63325  
probable membrane protein YNL339c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein N0152  
C;Species: Saccharomyces cerevisiae  
C;Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 09-Jul-2004  
C;Accession: S63325; S63326  
R;Obermaier, B.; Piravandi, E.; Rinke, M.  
submitted to the Protein Sequence Database, April 1996  
A;Reference number: S63317  
A;Accession: S63325  
A;Molecule type: DNA  
A;Residues: 1-916 <OBE>  
A;Cross-references: UNIPROT:P53819; EMBL:Z71615; MIPS:YNL339c  
A;Experimental source: strain S288C  
R;Wedler, H.; Wambutt, R.  
submitted to the Protein Sequence Database, April 1996  
A;Reference number: S63326  
A;Accession: S63326  
A;Molecule type: DNA  
A;Residues: 8-1859 <WED>  
A;Cross-references: EMBL:Z71615; MIPS:YNL339c  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:YRF1-6  
A;Cross-references: SGD:S0005283  
A;Map position: 14L  
A;Introns: 7/1  
A;Note: Closely related hypothetical proteins containing similar 12-residue repeats are

C;Keywords: tandem repeat; transmembrane protein  
F;8-24/Domain: transmembrane #status predicted <TMI>  
F;900-916/Domain: transmembrane #status predicted <TM2>  
F;1333-1452/Region: 12-residue repeats (T-T-T-[EKA]) - [SG] - [TI] - [ND] - [ASV] - [TRSNNK] - T- [NS] -  
Query Match 40.2%; Score 45; DB 2; Length 1859;  
Best Local Similarity 58.8%; Pred. No. 3.5e+02;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 4 YKLYNNKSSNSTLKNL 20  
Db 603 YTDYNNKSSSPRLKTL 619  
RESULT 31  
S64633  
probable membrane protein YGR296w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein G9608; hypothetical protein P0254; hypothetical  
C;Species: Saccharomyces cerevisiae  
C;Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 09-Jul-2004  
C;Accession: S64633; S64634; S65338; S65337  
R;Watson, M.D.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64630  
A;Accession: S64633  
A;Molecule type: DNA  
A;Residues: 1-1052 <WAT>  
A;Cross-references: UNIPROT:P53345; EMBL:Z73081; MIPS:YGR296w  
A;Experimental source: strain S288C  
A;Genetics: CH7  
R;Wedler, H.; Scharfe, M.; Wedler, E.; Wambutt, R.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64356  
A;Accession: S64634  
A;Molecule type: DNA  
A;Residues: 586-1859 <WED>  
A;Cross-references: EMBL:Z73081; MIPS:YGR296w  
A;Experimental source: strain S288C  
A;Genetics: CH7  
R;Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64792  
A;Accession: S65338  
A;Molecule type: DNA  
A;Residues: 1-1859 <WES>  
A;Cross-references: EMBL:Z73521; MIPS:YPL283c; NID:g1370581; PID:e246989; PID:g1370582  
A;Experimental source: strain S288C (AB972)  
A;Genetics: CH16  
R;Dellius, H.; Hebling, U.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64967  
A;Accession: S65337  
A;Molecule type: DNA  
A;Residues: 1-813 <DEL>  
A;Cross-references: EMBL:Z73521; MIPS:YPL283c  
A;Experimental source: strain S288C (AB972)  
A;Genetics: CH16  
C;Genetics: <CH7>  
A;Map position: 7R  
A;Introns: 7/1  
C;Genetics: <CH16>  
A;Map position: 16L  
A;Introns: 7/1  
C;Keywords: tandem repeat; transmembrane protein  
F;8-24/Domain: transmembrane #status predicted <TMI>  
F;900-916/Domain: transmembrane #status predicted <TM2>  
F;1333-1452/Region: 12-residue repeats (T-T-T-[EKA]) - [SG] - [TI] - [ND] - [ASV] - [TRSNNK] - T- [NS] -  
Query Match 40.2%; Score 45; DB 2; Length 1859;  
Best Local Similarity 58.8%; Pred. No. 3.5e+02;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 4 YKLYNNKSSNSTLKNL 20

Db 603 YTDYKNGSSPRLKTL 619  
RESULT 32  
A:54080  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 09-Jul-2004  
C:Accession: A54080: I50592  
R:Pang, K.S.; Barker, K.; Sudol, M.; Hanafusa, H.  
J. Biol. Chem. 269, 14056-14063, 1994  
A:Title: A transmembrane protein-tyrosine phosphatase contains spectrin-like repeats in  
A:Reference number: A54080; MUID:94245724; PMID:818686  
A:Accession: A54080  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1237 <FAN>  
A:Cross-references: UNIPROT:Q91976; EMBL:Z21960; NID:G510510; PIDN:CAA79972.1; PID:G510510  
C:Superfamily: leukocyte common antigen; leukocyte common antigen cytosolic domain homolog  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
F:528-1170/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:610-834/Domain: protein-tyrosine-phosphatase homology <PPP>  
F:786/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:792/Binding site: substrate phosphate (Arg) #status predicted  
Query Match 39.7%; Score 44.5; DB 2; Length 1237;  
Best Local Similarity 52.4%; Pred. No. 2.7e+02;  
Matches 11; Conservative 3; Mismatches 6; Indels 1; Gaps 1;  
Qy 3 DY-KLYNKNSSNSTLKNLGE 22  
Db 151 DYGNIETKNGSAEVLTKNLKE 171  
RESULT 33  
C75299  
tellurium resistance protein TerD - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: C75299  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75299  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-191 <WHI>  
A:Cross-references: UNIPROT:Q9RSA5; GB:AE002055; GB:AE000513; NID:G6460024; PIDN:AAF1176  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2221  
A:Map position: 1  
C:Superfamily: tellurium resistance protein terD  
Query Match 39.3%; Score 44; DB 2; Length 191;  
Best Local Similarity 33.3%; Pred. NO. 46;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 58 SDFIFYNKTSDGSGVEHTGD 78  
RESULT 34  
H97060  
uncharacterized conserved protein, probable metal-binding CAC1304 [imported] - Clostridi  
C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: H97060

R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: H97060  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-252 <KUR>  
A:Cross-references: UNIPROT:Q97JH9; GB:AE001437; PIDN:AAK79275.1; PID:GI5024234; GSPDB:G1  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1304  
Query Match 39.3%; Score 44; DB 2; Length 252;  
Best Local Similarity 42.1%; Pred. No. 61;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
Qy 2 SDYKLYNKNSSNSTLKNL 20  
Db 79 NEYKLYNKAGSNLNMIEKL 97  
RESULT 35  
T24854  
hypothetical protein T12A7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24854  
R:Lennard, N.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19943  
A:Accession: T24854  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-262 <WIL>  
A:Cross-references: UNIPROT:Q22421; EMBL:Z73911; PIDN:CAA98140.1; GSPDB:GN00022; CESP:T12  
C:Genetics:  
A:Gene: CESP:T12A7.3  
A:Map position: 4  
A:Introns: 86/2; 192/3; 231/2  
Query Match 39.3%; Score 44; DB 2; Length 262;  
Best Local Similarity 40.9%; Pred. No. 64;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 RSDYKLYNKNSSNSTLKNLGE 22  
Db 119 RSAYSFAQKNPSQHSFIPDIGE 140  
RESULT 36  
T32211  
hypothetical protein T03D3.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32211  
R:Murray, J.; Wohlmann, P.; Bauer, C.; Biewald, T.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid T03D3.  
A:Reference number: Z21136  
A:Accession: T32211  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-312 <MUR>  
A:Cross-references: EMBL:AF022980; PIDN:AAB69913.1; GSPDB:GN00023; CESP:T03D3.7  
A:Experimental source: strain Bristol N2; clone T03D3  
C:Genetics:  
A:Gene: CESP:T03D3.7  
A:Map position: 5  
A:Introns: 63/3; 186/3; 202/1; 258/2



Query Match 39.3%; Score 44; DB 2; Length 312;  
Best Local Similarity 66.7%; Pred. No. 76;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 KLYNKNSSNST 16  
|:|||||:|  
DB 299 KIYNKNTSSIT 310

## RESULT 37

T23572

hypothetical protein K10D3.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T23572

R;McMurray, A.

submitted to the EMBL Data Library, June 1996

A;Reference number: 219762

A;Accession: T23572

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-327 &lt;WIL&gt;

A;Cross-references: UNIPROT:Q21417; EMBL:Z75545; PIDN:CAA99885.1; GSPDB:GN00019; CESP:K1

A;Experimental source: clone K10D3

C;Genetics:

A;Gene: CESP:K10D3.3

A;Map position: 1

A;Introns: 46/2; 83/2; 169/3; 241/2

Query Match 39.3%; Score 44; DB 2; Length 327;  
Best Local Similarity 47.1%; Pred. No. 80;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 YKLYNKNSSNSTLKNL 20  
|:|:|:|:|  
DB 236 YEYVKSSPKSTIKRL 252

## RESULT 38

A90525

peptide chain release factor 1 (rf-1) [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C;Species: Mycoplasma pulmonis

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004

C;Accession: A90525

R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A;Reference number: A99512; MUID:21267165; PMID:11353084

A;Accession: A90525

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-359 &lt;KUR&gt;

A;Cross-references: UNIPROT:Q98RA5; GB:AL445566; PID:gi4089518; PIDN:CAC13278.1; GSPDB:C

A;Experimental source: strain UAB CTIP

C;Genetics:

A;Gene: MYPU 1050

A;Genetic code: SGC3

C;Superfamily: translation releasing factor

Query Match 39.3%; Score 44; DB 2; Length 359;  
Best Local Similarity 42.9%; Pred. No. 88;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDYKLYNKNSSNSTLKNLGE 22  
|:|:|:|:|:|:|:|  
DB 30 SDIKLYKFSRELNSIKISE 50

## RESULT 39

A23535

clustered asparagine-rich merozoite-associated antigen - malaria parasite (Plasmodium fa

C;Species: Plasmodium falciparum

C;Date: 02-Jun-1988 #sequence\_revision 02-Jun-1988 #text\_change 09-Jul-2004

C;Accession: A23535; A49245

R;Wahlgren, M.; Aslund, L.; Franzen, L.; Sundvall, M.; Wahlin, B.; Berzins, K.; McNicol,

Proc. Natl. Acad. Sci. U.S.A. 83, 2677-2681, 1986

A;Title: A Plasmodium falciparum antigen containing clusters of asparagine residues.

A;Reference number: A23535; MUID:86206015; PMID:3517875

A;Accession: A23535

A;Molecule type: DNA

A;Residues: 1-451 &lt;WAH&gt;

A;Cross-references: UNIPROT:P13824; GB:M13021; NID:gi160079; PIDN:AAA29485.1; PID:9552179

R;Sjlander, A.; Stahl, S.; Lovgren, K.; Hansson, M.; Cavellier, L.; Waller, A.; Helmbj,

Exp. Parasitol. 76, 134-145, 1993

A;Title: Plasmodium falciparum: the immune response in rabbits to the clustered asparagi

A;Reference number: A49245; MUID:93202225; PMID:8454022

A;Contents: 768

A;Accession: A49245

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 9-25 &lt;SJO&gt;

A;Cross-references: GB:S56699; NID:q298738; PID:q298739

A;Note: sequence extracted from NCBI backbone (NCBIN:127859, NCBI:P127860)

Query Match 39.3%; Score 44; DB 2; Length 451;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 SDYKLYNKNSSNS 15  
|:|:|:|:|:|:|  
DB 164 NNTNFYNNSSNN 177

## RESULT 40

T52382

zinc finger protein ZPT4-4, C2H2-type [imported] - garden petunia

C;Species: Petunia x hybrida (garden petunia)

C;Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 09-Jul-2004

C;Accession: T52382

R;Kubo, K.; Sakamoto, A.; Kobayashi, A.; Rybka, Z.; Kanno, Y.; Nakagawa, H.; Nishino, T.

Nucleic Acids Res. 26, 608-615, 1998

A;Title: Cys2/His2 zinc-finger protein family of petunia: evolution and general mechanis

A;Reference number: Z26061

A;Accession: T52382

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-477 &lt;KUB&gt;

A;Cross-references: UNIPROT:O22091; EMBL:AB006606; PIDN:BAA21928.1

A;Experimental source: strain Mitchell diploid

Query Match 39.3%; Score 44; DB 2; Length 477;  
Best Local Similarity 38.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 SDYKLYNKNSSNSTLKNLGE 22  
|:|:|:|:|:|:|  
DB 357 TDSKLIKNSSKNSTIDFGE 377

Search completed: November 24, 2004, 09:30:15  
Job time : 21.1739 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:29:11 ; Search time 82.9783 Seconds  
(without alignments)  
94.040 Million cell updates/sec

Title: US-09-719-379a-2

Perfect score: 112

Sequence: 1 RSDYKLYNKNSSNSTLKNLGE 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                      | Description        |
|------------|-------|-------------|--------|-------------------------|--------------------|
| 1          | 112   | 100.0       | 22     | 14 US-10-203-942-2      | Sequence 2, Appli  |
| 2          | 112   | 100.0       | 31     | 14 US-10-203-942-6      | Sequence 6, Appli  |
| 3          | 81.5  | 72.8        | 19     | 14 US-10-203-942-4      | Sequence 4, Appli  |
| 4          | 81.5  | 72.8        | 28     | 14 US-10-203-942-8      | Sequence 8, Appli  |
| 5          | 81.5  | 72.8        | 353    | 16 US-10-467-421-21     | Sequence 21, Appli |
| 6          | 76.5  | 68.3        | 352    | 14 US-10-336-840-37     | Sequence 37, Appli |
| 7          | 76.5  | 68.3        | 353    | 14 US-10-203-942-9      | Sequence 9, Appli  |
| 8          | 52    | 46.4        | 322    | 14 US-10-156-761-12985  | Sequence 12985, A  |
| 9          | 50    | 44.6        | 101    | 15 US-10-424-599-202105 | Sequence 202105,   |
| 10         | 50    | 44.6        | 141    | 16 US-10-437-963-142742 | Sequence 142742,   |
| 11         | 48    | 42.9        | 172    | 16 US-10-437-963-123218 | Sequence 123218,   |
| 12         | 48    | 42.9        | 706    | 15 US-10-282-122A-78118 | Sequence 78118, A  |
| 13         | 47    | 42.0        | 198    | 15 US-10-424-599-208738 | Sequence 208738,   |

|    |      |      |      |    |                      |                    |
|----|------|------|------|----|----------------------|--------------------|
| 14 | 47   | 42.0 | 764  | 15 | US-10-424-599-143917 | Sequence 143917,   |
| 15 | 46.5 | 41.5 | 271  | 15 | US-10-424-599-238326 | Sequence 238326,   |
| 16 | 46.5 | 41.5 | 271  | 17 | US-10-739-930-9182   | Sequence 9182, Ap  |
| 17 | 46   | 41.1 | 51   | 16 | US-10-767-701-61375  | Sequence 61375, A  |
| 18 | 46   | 41.1 | 652  | 16 | US-10-437-963-172725 | Sequence 172725,   |
| 19 | 46   | 41.1 | 2052 | 15 | US-10-282-122A-51602 | Sequence 51602, A  |
| 20 | 45   | 40.2 | 70   | 17 | US-10-425-115-305355 | Sequence 305355,   |
| 21 | 45   | 40.2 | 71   | 17 | US-10-425-115-297730 | Sequence 297730,   |
| 22 | 45   | 40.2 | 182  | 14 | US-10-369-493-6144   | Sequence 6144, Ap  |
| 23 | 45   | 40.2 | 220  | 15 | US-10-424-599-188660 | Sequence 188660,   |
| 24 | 45   | 40.2 | 225  | 16 | US-10-607-559-54     | Sequence 54, Appl  |
| 25 | 45   | 40.2 | 338  | 16 | US-10-451-467A-44    | Sequence 44, Appl  |
| 26 | 45   | 40.2 | 589  | 15 | US-10-335-977-6022   | Sequence 6022, Ap  |
| 27 | 45   | 40.2 | 593  | 15 | US-10-335-977-6023   | Sequence 6023, Ap  |
| 28 | 45   | 40.2 | 1796 | 16 | US-10-451-467A-138   | Sequence 138, App  |
| 29 | 45   | 40.2 | 1841 | 14 | US-10-083-357-1341   | Sequence 1341, Ap  |
| 30 | 44.5 | 39.7 | 92   | 17 | US-10-425-115-279956 | Sequence 279956,   |
| 31 | 44   | 39.3 | 51   | 15 | US-10-424-599-272204 | Sequence 272204,   |
| 32 | 44   | 39.3 | 101  | 16 | US-10-437-963-126777 | Sequence 126777,   |
| 33 | 44   | 39.3 | 127  | 14 | US-10-104-047-3149   | Sequence 3149, Ap  |
| 34 | 44   | 39.3 | 159  | 16 | US-10-767-701-62083  | Sequence 62083, A  |
| 35 | 44   | 39.3 | 159  | 17 | US-10-425-115-221399 | Sequence 221399,   |
| 36 | 44   | 39.3 | 200  | 16 | US-10-437-963-176751 | Sequence 176751,   |
| 37 | 44   | 39.3 | 256  | 16 | US-10-437-963-116447 | Sequence 116447,   |
| 38 | 44   | 39.3 | 368  | 15 | US-10-469-993-12     | Sequence 12, Appl  |
| 39 | 44   | 39.3 | 369  | 11 | US-09-809-665A-153   | Sequence 153, App  |
| 40 | 44   | 39.3 | 529  | 15 | US-10-149-310-240    | Sequence 240, App  |
| 41 | 44   | 39.3 | 1324 | 14 | US-10-369-493-3227   | Sequence 3227, App |
| 42 | 43.5 | 38.8 | 46   | 15 | US-10-424-599-146089 | Sequence 146089,   |
| 43 | 43.5 | 38.8 | 166  | 15 | US-10-282-122A-52791 | Sequence 52791, A  |
| 44 | 43.5 | 38.8 | 553  | 9  | US-09-738-626-5482   | Sequence 5482, Ap  |
| 45 | 43   | 38.4 | 86   | 9  | US-09-864-761-46557  | Sequence 46557, A  |

#### ALIGNMENTS

RESULT 1  
US-10-203-942-2  
; Sequence 2, Application US/10203942  
; Publication No. US20030096370A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: DENOSL, PHILIPPE  
; APPLICANT: POOLMAN, JAN  
; APPLICANT: THONNARD, JOELLE  
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN AND USE THEREOF IN VACCINATION  
; FILE REFERENCE: B45210  
; CURRENT APPLICATION NUMBER: US/10/203,942  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: PCT/EP01/01556  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: GB 0003502.2  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-203-942-2

Query Match 100.0%; Score 112; DB 14; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RSDYKLYNKNSSNSTLKNLGE 22

Db 1 RSDYKLYNKNSSNSTLKNLGE 22

RESULT 2

```
US-10-203-942-6
; Sequence 6, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-6
Query Match 100.0%; Score 112; DB 14; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 3 RSDYKLYNKSSNSTLKNLGE 24
RESULT 3
US-10-203-942-4
; Sequence 4, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-4
Query Match 72.8%; Score 81.5; DB 14; Length 19;
Best Local Similarity 81.8%; Pred. No. 3.7e-05;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 1 RSDYKLYNKNS---STLKDILGE 19
RESULT 4
US-10-203-942-8
; Sequence 8, Application US/10203942
; Publication No. US20030096370A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: DENOEL, PHILIPPE
; APPLICANT: POOLMAN, JAN
; APPLICANT: THONNARD, JOELLE
; TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTER MEMBRANE
; FILE REFERENCE: B45210
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: PCT/EP01/01556
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: GB 0003502.2
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-203-942-8
Query Match 72.8%; Score 81.5; DB 14; Length 28;
Best Local Similarity 81.8%; Pred. No. 5.7e-05;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 3 RSDYKLYNKNS---STLKDILGE 21
RESULT 5
US-10-467-421-21
; Sequence 21, Application US/10467421
; Publication No. US20040116665A1
; GENERAL INFORMATION:
; APPLICANT: BERTHET, Francois-Xavier Jacques
; APPLICANT: DENOEL, Philippe
; APPLICANT: NEYT, Cecile Anne
; APPLICANT: POOLMAN, Jan
; APPLICANT: THONNARD, Joelle
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45259
; CURRENT APPLICATION NUMBER: US/10/467,421
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01361
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103171.5
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; NAME/KEY: VARIANT
; LOCATION: (1)...(353)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-467-421-21
Query Match 72.8%; Score 81.5; DB 16; Length 353;
Best Local Similarity 81.8%; Pred. No. 0.00099;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 132 RSDYKLYNKNS---STLKDILGE 150
RESULT 6
US-10-336-840-37
; Sequence 37, Application US/10336840
```

RESULT 8  
US-10-156-761-12985  
; Sequence 12985, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN

```

RESULT 10
US-10-437-963-142742
; Sequence 142742, Application US/10437963
; Publication No. US20040123343A1
;
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
;

```

; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 142742  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_43719C.1.pep  
US-10-437-963-142742

Query Match 44.6%; Score 50; DB 16; Length 141;  
Best Local Similarity 57.9%; Pred. No. 16;  
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKN 19  
Db 21 RDVYWCYNKNKRISTLKN 39

RESULT 11  
US-10-437-963-123218  
; Sequence 123218, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 123218  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_26074C.1.pep  
US-10-437-963-123218

Query Match 42.9%; Score 48; DB 16; Length 172;  
Best Local Similarity 40.9%; Pred. No. 40;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKN 22  
Db 7 RKVFMFDKNGDGRITKELGE 28

RESULT 12  
US-10-282-122A-78118  
; Sequence 78118, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78118  
; LENGTH: 706  
; TYPE: PRT  
; ORGANISM: Versinia pestis  
US-10-282-122A-78118

Query Match 42.9%; Score 48; DB 15; Length 706;  
Best Local Similarity 40.0%; Pred. No. 2e+02;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSNSTLKNLGE 22  
Db 543 DYERFNVNKRASNTVLSLNQ 562

RESULT 13  
US-10-424-599-208738  
; Sequence 208738, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 208738  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_30519C.1.pep  
US-10-424-599-208738

Query Match 42.0%; Score 47; DB 15; Length 198;

```
Best Local Similarity 64.3%; Pred. No. 66;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 9 KNSSNSTLKNLGE 22
   |||:|:|:|:|
Db 175 KNTSQSSKNLGE 188

RESULT 14
US-10-424-599-143917
; Sequence 143917, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 143917
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_100970C.1.pep
US-10-424-599-143917

Query Match 42.0%; Score 47; DB 15; Length 764;
Best Local Similarity 56.2%; Pred. No. 3e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 7 YKNSSNSTLKNLGE 22
   |||:|:|:|:|
Db 714 WNRGSSNTTSKMGQ 729

RESULT 15
US-10-424-599-238326
; Sequence 238326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 238326
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_57233C.1.pep
US-10-424-599-238326

Query Match 41.5%; Score 46.5; DB 15; Length 271;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 12; Conservative 1; Mismatches 5; Indels 9; Gaps 1;

Qy 4 YKLYKNSSS-----NSTLKNLG 21
   |||:|:|:|
Db 3 YSFYDKTSSSGSDISGLFNPTLNLG 29

RESULT 16
US-10-424-599-238326
; Sequence 238326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 238326
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_57233C.1.pep
US-10-424-599-238326

Query Match 41.5%; Score 46.5; DB 15; Length 271;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 12; Conservative 1; Mismatches 5; Indels 9; Gaps 1;

Qy 4 YKLYKNSSS-----NSTLKNLG 21
   |||:|:|:|
Db 3 YSFYDKTSSSGSDISGLFNPTLNLG 29

RESULT 17
US-10-767-701-61375
; Sequence 61375, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 61375
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9305245.pep
US-10-767-701-61375

Query Match 41.1%; Score 46; DB 16; Length 51;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 6; Mismatches 4; Indels 2; Gaps 1;

Qy 2 SDYKLYKNSS--SNSTLKNLG 21
   |||:|:|:|
Db 13 SPYKGYNRNSTLFFNNIAKSIG 34

RESULT 18
US-10-437-963-172725
; Sequence 172725, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```
US-10-739-930-9182
; Sequence 9182, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 9182
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-C49987_1.p
US-10-739-930-9182

Query Match 41.5%; Score 46.5; DB 17; Length 271;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 12; Conservative 1; Mismatches 5; Indels 9; Gaps 1;

Qy 4 YKLYKNSSS-----NSTLKNLG 21
   |||:|:|:|
Db 3 YSFYDKTSSSGSDISGLFNPTLNLG 29

RESULT 17
US-10-767-701-61375
; Sequence 61375, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 61375
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9305245.pep
US-10-767-701-61375

Query Match 41.1%; Score 46; DB 16; Length 51;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 6; Mismatches 4; Indels 2; Gaps 1;

Qy 2 SDYKLYKNSS--SNSTLKNLG 21
   |||:|:|:|
Db 13 SPYKGYNRNSTLFFNNIAKSIG 34

RESULT 18
US-10-437-963-172725
; Sequence 172725, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 172725  
; LENGTH: 652  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_70833C.1.pep  
US-10-437-963-172725

Query Match 41.1%; Score 46; DB 16; Length 652;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DYKLYNKNSSNS 15  
||:|||||:  
Db 169 DYFVNANSSSS 181

RESULT 19  
US-10-282-122A-51602  
; Sequence 51602, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51602  
; LENGTH: 2052  
; TYPE: PRT  
; ORGANISM: Clostridium acetobutylicum  
US-10-282-122A-51602

Query Match 41.1%; Score 46; DB 15; Length 2052;

Best Local Similarity 55.6%; Pred. No. 1.3e+03;  
Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 1;  
Qy 7 YNKNSS--SNSTLKNLGE 22  
:|||||:|||||:  
Db 1646 FEKNSSNFSKNTLKNIGD 1663

## RESULT 20

US-10-425-115-305355  
; Sequence 305355, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 305355  
; LENGTH: 70  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_41553C.1.pep  
US-10-425-115-305355

Query Match 40.2%; Score 45; DB 17; Length 70;  
Best Local Similarity 69.2%; Pred. No. 40;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 YNKNSSSNSTLKN 19  
||:|||||:  
Db 3 YSKHSSRRSTLKN 15

## RESULT 21

US-10-425-115-297730  
; Sequence 297730, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 297730  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(71)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_34603C.1.pep  
US-10-425-115-297730

Query Match 40.2%; Score 45; DB 17; Length 71;  
Best Local Similarity 64.3%; Pred. No. 41;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 YNKNSSSNSTLKNL 20  
:|||||:|||||:





Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6022:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...589  
SEQUENCE DESCRIPTION: SEQ ID NO: 6022:  
US-10-335-977-6022  
Query Match 40.2%; Score 45; DB 15; Length 589;  
Best Local Similarity 58.8%; Pred. No. 4.5e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
Qy 4 YKLYNKSSNSTLKNL 20  
Db 376 YALVGKNASGKSTLINL 392  
RESULT 27  
US-10-335-977-6023  
Sequence 6023, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6023:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 593 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...593  
SEQUENCE DESCRIPTION: SEQ ID NO: 6023:  
US-10-335-977-6023  
Query Match 40.2%; Score 45; DB 15; Length 593;  
Best Local Similarity 58.8%; Pred. No. 4.5e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
Qy 4 YKLYNKSSNSTLKNL 20  
Db 380 YALVGKNASGKSTLINL 396  
RESULT 28  
US-10-451-467A-138  
Sequence 138, Application US/10451467A  
Publication No. US20040161840A1  
GENERAL INFORMATION:  
APPLICANT: CONTRERAS, ROLAND HENRI  
APPLICANT: EBERHARDT, INES  
APPLICANT: LUYTEN, WALTER HERMAN MARIA LOUIS  
APPLICANT: REEKMAN, RIEKA JOSEPHINA  
TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN  
TITLE OF INVENTION: YEAST AND FUNGI  
FILE REFERENCE: JAB-1667  
CURRENT APPLICATION NUMBER: US/10/451,467A  
CURRENT FILING DATE: 2003-06-19  
PRIOR APPLICATION NUMBER: EP 00870318.3  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: EP 01870002.1  
PRIOR FILING DATE: 2001-01-04  
PRIOR APPLICATION NUMBER: EP 01870003.9  
PRIOR FILING DATE: 2001-01-09  
NUMBER OF SEQ ID NOS: 732  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 138  
LENGTH: 1796  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-451-467A-138  
Query Match 40.2%; Score 45; DB 16; Length 1796;  
Best Local Similarity 58.8%; Pred. No. 1.6e+03;  
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy 4 YKLYNKSSNSTLKNL 20
Db 539 YTDYKNGSSPRKLT 555

RESULT 29
US-10-083-357-1341
; Sequence 1341, Application US/10083357
; Publication No. US20030054370A1
; GENERAL INFORMATION:
; APPLICANT: Qiangdong Zeng et al.
; TITLE OF INVENTION: Systemic Discovery of New Genes
; FILE REFERENCE: 032796-090
; CURRENT APPLICATION NUMBER: US/10/083.357
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 1346
; SEQ ID NO 1341
; LENGTH: 1841
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1341

Query Match 40.2%; Score 45; DB 14; Length 1841;
Best Local Similarity 58.8%; Pred. No. 1.6e+03;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 YKLYNKSSNSTLKNL 20
Db 584 YTDYKNGSSPRKLT 600

RESULT 30
US-10-425-115-279956
; Sequence 279956, Application US/10425115
; Publication No. US2004021427A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279956
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_186899C.1.pep
US-10-425-115-279956

Query Match 39.7%; Score 44.5; DB 17; Length 92;
Best Local Similarity 43.5%; Pred. No. 65;
Matches 10; Conservative 4; Mismatches 4; Indels 5; Gaps 1;

Qy 4 YKLYNKSSNSTL-----KNLG 21
Db 45 FTTYNHASNTNLDHNTKRLG 67

RESULT 31
US-10-424-599-272204
; Sequence 272204, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 272204
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(51)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87819C.1.pep
US-10-424-599-272204

Query Match 39.3%; Score 44; DB 15; Length 51;
Best Local Similarity 64.3%; Pred. No. 40;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 9 KNSSNSTLKNLGE 22
Db 15 ENLTKNSFLKNLGE 28

RESULT 32
US-10-437-963-126777
; Sequence 126777, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 126777
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_29291C.1.pep
US-10-437-963-126777

Query Match 39.3%; Score 44; DB 16; Length 101;
Best Local Similarity 42.9%; Pred. No. 86;
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDVLYNKSSNSTLKNLGE 22
Db 75 SQLKLTLDKSGKSTIKDMGK 95

RESULT 33
US-10-104-047-3149
; Sequence 3149, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104.047
; CURRENT FILING DATE: 2002-03-25
```

; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3149  
; LENGTH: 127  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3149

Query Match 39.3%; Score 44; DB 14; Length 127;  
Best Local Similarity 47.6%; Pred. No. 1.1e+02;  
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 31 TNVRNVTKLSVSNQTLKDIGE 51

## RESULT 34

US-10-767-701-62083  
; Sequence 62083, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53355)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 62083  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 9855880.pcp  
US-10-767-701-62083

Query Match 39.3%; Score 44; DB 16; Length 159;  
Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 8 NKNSSNSTLKNLGE 22  
Db 59 NRNSSSSASRRKLGE 73

## RESULT 35

US-10-425-115-221399  
; Sequence 221399, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 221399  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_133507C.1.pcp  
US-10-425-115-221399

Query Match 39.3%; Score 44; DB 17; Length 159;  
Best Local Similarity 36.4%; Pred. No. 1.4e+02;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 RSDYKLYNKNSSNSTLKNLGE 22  
Db 10 RKVFQMFDRKNGDGOITKKELGE 31

## RESULT 36

US-10-437-963-176751  
; Sequence 176751, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 176751  
; LENGTH: 200  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(200)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_74470C.1.pcp  
US-10-437-963-176751

Query Match 39.3%; Score 44; DB 16; Length 200;  
Best Local Similarity 42.9%; Pred. No. 1.9e+02;  
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 37 NDAKLDKNDSDNNALVKHLEE 57

## RESULT 37

US-10-437-963-116447  
; Sequence 116447, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 116447  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:

```
; NAME/KEY: unsure
; LOCATION: (1)..(256)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19948C.1.pap
US-10-437-963-116447

Query Match      39.3%; Score 44; DB 16; Length 256;
Best Local Similarity 60.0%; Pred. NO. 2.5e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy      8 NKNSSNSTLKNLGE 22
Db      4 DKSSSTGTLKNLQE 18

RESULT 38
US-10-469-993-12
; Sequence 12, Application US/10469993
; Publication No. US20040078847A1
; GENERAL INFORMATION:
; APPLICANT: Paldi, Nitzan
; TITLE OF INVENTION: METHOD OF ENHANCING ENTOMOPHILOUS
; FILE REFERENCE: 26678
; CURRENT APPLICATION NUMBER: US/10/469,993
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Clarkia breweri
US-10-469-993-12

Query Match      39.3%; Score 44; DB 15; Length 368;
Best Local Similarity 61.5%; Pred. NO. 3.7e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      5 KLYNKNSSNSTL 17
Db     177 KVFNKGMSNSTI 189

RESULT 39
US-09-809-665A-153
; Sequence 153, Application US/09809665A
; Publication No. US20040110268A1
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Actinobacillus pleuropneumoniae
US-09-809-665A-153

Query Match      39.3%; Score 44; DB 11; Length 369;
Best Local Similarity 44.4%; Pred. NO. 3.7e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy      1 RSDYKLYNKNSSNSTLK 18
Db     177 KVFNKGMSNSTI 189
```

```
Db     139 RNDYKKYGAENTNESTTK 156

RESULT 40
US-10-149-310-240
; Sequence 240, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Madden, Kevin T.
; APPLICANT: Maxon, Mary
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 240
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-149-310-240

Query Match      39.3%; Score 44; DB 15; Length 529;
Best Local Similarity 41.2%; Pred. NO. 5.6e+02;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy      6 LYNNSSNSTLKNLGE 22
Db     104 IYNNSSNSTLNVNNGE 120

Search completed: November 24, 2004, 10:00:13
Job time : 84.9783 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:14:14 ; Search time 19.3696 Seconds  
(without alignments)  
75.324 Million cell updates/sec

Title: US-09-719-379A-2

Perfect score: 112

Sequence: 1 RSDYKLYNKSSNSTLKNLGE 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 50    | 44.6        | 219    | 4     | US-09-270-767-58822  |
| 2          | 47    | 42.0        | 482    | 4     | US-09-270-767-43292  |
| 3          | 47    | 42.0        | 820    | 4     | US-09-248-796A-17231 |
| 4          | 46    | 41.1        | 135    | 4     | US-09-248-796A-27899 |
| 5          | 46    | 41.1        | 179    | 4     | US-09-270-767-37611  |
| 6          | 46    | 41.1        | 179    | 4     | US-09-270-767-52828  |
| 7          | 46    | 41.1        | 300    | 4     | US-09-270-767-44496  |
| 8          | 45    | 40.2        | 60     | 4     | US-09-543-681A-4823  |
| 9          | 45    | 40.2        | 159    | 4     | US-09-248-796A-21631 |
| 10         | 45    | 40.2        | 225    | 4     | US-09-645-055-54     |
| 11         | 45    | 40.2        | 249    | 4     | US-09-248-796A-15815 |
| 12         | 45    | 40.2        | 733    | 4     | US-09-328-352-5599   |
| 13         | 45    | 40.2        | 1382   | 4     | US-09-538-092-590    |
| 14         | 44.5  | 39.7        | 1237   | 1     | US-08-241-853-2      |
| 15         | 44.5  | 39.7        | 1237   | 2     | US-08-850-917-2      |
| 16         | 44    | 39.3        | 72     | 4     | US-09-248-796A-26540 |
| 17         | 44    | 39.3        | 213    | 4     | US-09-270-767-32157  |
| 18         | 44    | 39.3        | 213    | 4     | US-09-270-767-47374  |
| 19         | 44    | 39.3        | 312    | 4     | US-09-248-796A-17634 |
| 20         | 44    | 39.3        | 369    | 4     | US-09-809-665A-153   |
| 21         | 43.5  | 38.8        | 120    | 4     | US-09-134-000C-4078  |
| 22         | 43.5  | 38.8        | 214    | 4     | US-09-134-000C-6579  |
| 23         | 43.5  | 38.8        | 214    | 4     | US-09-134-000C-6679  |
| 24         | 43    | 38.4        | 78     | 4     | US-09-270-767-34030  |
| 25         | 43    | 38.4        | 78     | 4     | US-09-270-767-49247  |
| 26         | 43    | 38.4        | 116    | 1     | US-08-690-102A-4     |
| 27         | 43    | 38.4        | 116    | 3     | US-09-127-902-4      |

|    |      |      |      |   |                      |                    |
|----|------|------|------|---|----------------------|--------------------|
| 28 | 43   | 38.4 | 116  | 3 | US-09-155-107-4      | Sequence 4, Appli  |
| 29 | 43   | 38.4 | 116  | 5 | PCT-US95-09641-4     | Sequence 4, Appli  |
| 30 | 43   | 38.4 | 169  | 5 | PCT-US96-07709-22    | Sequence 22, Appli |
| 31 | 43   | 38.4 | 185  | 4 | US-09-583-110-3991   | Sequence 3991, Ap  |
| 32 | 43   | 38.4 | 195  | 4 | US-09-270-767-60029  | Sequence 60029, A  |
| 33 | 43   | 38.4 | 197  | 5 | PCT-US96-07709-19    | Sequence 19, Appli |
| 34 | 43   | 38.4 | 204  | 5 | PCT-US96-07709-33    | Sequence 33, Appli |
| 35 | 43   | 38.4 | 220  | 5 | PCT-US96-07709-30    | Sequence 30, Appli |
| 36 | 43   | 38.4 | 223  | 4 | US-09-541-759-1      | Sequence 1, Appli  |
| 37 | 43   | 38.4 | 244  | 4 | US-09-270-767-44586  | Sequence 44586, A  |
| 38 | 43   | 38.4 | 248  | 5 | PCT-US96-07709-25    | Sequence 25, Appli |
| 39 | 43   | 38.4 | 281  | 4 | US-09-134-000C-4344  | Sequence 4344, Ap  |
| 40 | 43   | 38.4 | 307  | 4 | US-09-270-767-44579  | Sequence 44579, A  |
| 41 | 43   | 38.4 | 683  | 3 | US-09-213-293D-1     | Sequence 1, Appli  |
| 42 | 43   | 38.4 | 866  | 4 | US-09-527-084A-4     | Sequence 4, Appli  |
| 43 | 43   | 38.4 | 1255 | 4 | US-09-248-796A-14158 | Sequence 14158, A  |
| 44 | 42.5 | 37.9 | 300  | 4 | US-09-543-681A-4416  | Sequence 4416, Ap  |
| 45 | 42   | 37.5 | 62   | 4 | US-09-248-796A-23261 | Sequence 23261, A  |

ALIGNMENTS

RESULT 1

US-09-270-767-58822  
; Sequence 58822, Application US/09270767  
; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 58822

; LENGTH: 219

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-58822

Query Match 44.6%; Score 50; DB 4; Length 219;

Best Local Similarity 57.9%; Pred. No. 4.3;

Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLKN 19

Db 101 RLNYKULFLCRNSQSSTLKN 119

RESULT 2

US-09-270-767-43292  
; Sequence 43292, Application US/09270767  
; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 43292

; LENGTH: 482

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-43292

Query Match 42.0%; Score 47; DB 4; Length 482;

Best Local Similarity 64.7%; Pred. No. 32;  
Matches 11; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy 2 SDYKLYNKSSNSTLTK 18  
Db 133 SDYNI--KNSSNPTLK 147

RESULT 3  
US-09-248-796A-17231  
; Sequence 17231, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17231  
; LENGTH: 820  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17231

Query Match 42.0%; Score 47; DB 4; Length 820;  
Best Local Similarity 42.9%; Pred. No. 60;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLKNLG 21  
Db 528 RKFKAIKRNEQSGKTLNLG 548

RESULT 4  
US-09-248-796A-27899  
; Sequence 27899, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 27899  
; LENGTH: 135  
; TYPE: PRT  
; ORGANISM: Candida albicans  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (133)  
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknd  
US-09-248-796A-27899

Query Match 41.1%; Score 46; DB 4; Length 135;  
Best Local Similarity 47.4%; Pred. No. 10;  
Matches 9; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSTLKNL 20  
Db 78 SNYSNYSNNSSNNTLKIL 96

RESULT 5  
US-09-270-767-37611  
; Sequence 37611, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 37611  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-37611

Query Match 41.1%; Score 46; DB 4; Length 179;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKLYNKSSNSTLKN 19  
Db 100 YNILNKSGDTNCTLKN 115

RESULT 6  
US-09-270-767-52828  
; Sequence 52828, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 52828  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-52828

Query Match 41.1%; Score 46; DB 4; Length 179;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKLYNKSSNSTLKN 19  
Db 100 YNILNKSGDTNCTLKN 115

RESULT 7  
US-09-270-767-44496  
; Sequence 44496, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44496  
; LENGTH: 300



```
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44496

Query Match      41.1%; Score 46; DB 4; Length 300;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNST 16
    ||| :||| :
Db 277 RSDLRLEFNKKKNT 292

RESULT 8
US-09-543-681A-4823
; Sequence 4823, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4823
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4823

Query Match      40.2%; Score 45; DB 4; Length 60;
Best Local Similarity 56.2%; Pred. No. 5.7;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 5 KLYNKSSNSTLKNL 20
    ||| :||| :
Db 3 KIYKNSITSTLKR 18

RESULT 9
US-09-248-796A-21631
; Sequence 21631, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21631
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21631

Query Match      40.2%; Score 45; DB 4; Length 159;
Best Local Similarity 42.1%; Pred. No. 18;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLKN 19
    ||| :||| :
Db 136 KQYKSHHTNNSNSVNN 154
```

```
RESULT 10
US-09-645-055-54
; Sequence 54, Application US/09645055
; Patent No. 6593701
; GENERAL INFORMATION:
; APPLICANT: CLARITY BIOSCIENCES, INC.
; APPLICANT: Honeycutt, Rhonda
; APPLICANT: McClelland, Michael
; TITLE OF INVENTION: IDENTIFYING ORGANISMS BY DETECTING
; FILE REFERENCE: 475402000100
; CURRENT APPLICATION NUMBER: US/09/645,055
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 60/150,977
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Tilletia indica
US-09-645-055-54

Query Match      40.2%; Score 45; DB 4; Length 225;
Best Local Similarity 47.4%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 4 YKLYNKSSNSTLKNLGE 22
    ||| :||| :||| :
Db 182 YKTYTKKASSYTTWKEVSE 200

RESULT 11
US-09-248-796A-15815
; Sequence 15815, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15815
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15815

Query Match      40.2%; Score 45; DB 4; Length 249;
Best Local Similarity 46.7%; Pred. No. 31;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 6 LYNNKSSNSTLKNL 20
    ||| :||| :||| :
Db 24 IYNNNSNSKTYKNI 38

RESULT 12
US-09-328-352-5599
; Sequence 5599, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
```



; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1237 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Gallus domesticus  
; STRAIN: DT40  
; US-08-850-917-2

Query Match 39.7%; Score 44.5; DB 2; Length 1237;  
Best Local Similarity 52.4%; Pred. No. 2.4e+02;  
Matches 11; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 3 DY-KLYNKNSSNSTLKNLGE 22  
Db 151 DYGNIEEKNSAEVTLKNLKE 171

RESULT 16  
US-09-248-796A-26540  
; Sequence 26540, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 26540  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-26540

Query Match 39.3%; Score 44; DB 4; Length 72;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNL 20  
Db 48 RSNFKLKLKNYPRNPTORNL 67

RESULT 17  
US-09-270-767-32157  
; Sequence 32157, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32157  
; LENGTH: 213  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-32157

Query Match 39.3%; Score 44; DB 4; Length 213;  
Best Local Similarity 38.1%; Pred. No. 36;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 8 TDYNIYNKYHSNNNFNKKTNQ 28

## RESULT 18

US-09-270-767-47374  
; Sequence 47374, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47374  
; LENGTH: 213  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-47374

Query Match 39.3%; Score 44; DB 4; Length 213;  
Best Local Similarity 38.1%; Pred. No. 36;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 2 SDYKLYNKNSSNSTLKNLGE 22  
Db 8 TDYNIYNKYHSNNNFNKKTNQ 28

## RESULT 19

US-09-248-796A-17634  
; Sequence 17634, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17634  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17634

Query Match 39.3%; Score 44; DB 4; Length 312;  
Best Local Similarity 44.4%; Pred. No. 57;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YKLYNKNSSNSTLKNLKG 21  
Db 248 HDLFFNNRNSNSVGKMG 265

## RESULT 20

US-09-809-665A-153  
; Sequence 153, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; CURRENT FILING DATE: 2001-03-15

```
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Actinobacillus pleuropneumoniae
US-09-809-865A-153

Query Match      39.3%; Score 44; DB 4; Length 369;
Best Local Similarity 44.4%; Pred. No. 70;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLK 18
Db 139 RNDYKKGAEANTNESTTK 156

RESULT 21
US-09-134-000C-4078
; Sequence 4078, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4078
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: Amino acid 15 is Xaa wherein Xaa = any amino acid.
US-09-134-000C-4078

Query Match      38.8%; Score 43.5; DB 4; Length 120;
Best Local Similarity 45.0%; Pred. No. 22;
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 4 YKL-YNKSSNSTLKNLGE 22
Db 98 YKVFNEESKKFTTLENLGE 117

RESULT 22
US-09-134-000C-6579
; Sequence 6579, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6579

Query Match      38.8%; Score 43.5; DB 4; Length 214;
Best Local Similarity 45.0%; Pred. No. 44;
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 4 YKL-YNKSSNSTLKNLGE 22
Db 192 YKVFNEESKKFTTLENLGE 211

RESULT 23
US-09-134-000C-6679
; Sequence 6679, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6679
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6679

Query Match      38.8%; Score 43.5; DB 4; Length 214;
Best Local Similarity 45.0%; Pred. No. 44;
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 4 YKL-YNKSSNSTLKNLGE 22
Db 192 YKVFNEESKKFTTLENLGE 211

RESULT 24
US-09-270-767-34030
; Sequence 34030, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34030
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-34030

Query Match      38.4%; Score 43; DB 4; Length 78;
Best Local Similarity 44.4%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYNKSSNSTLKN 19
Db 44 NNYKHLTTTSSNSTASN 61
```

## RESULT 25

US-09-270-767-49247  
; Sequence 49247, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49247  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-49247

Query Match 38.4%; Score 43; DB 4; Length 78;  
Best Local Similarity 44.4%; Pred. No. 16;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDYKLYKNSSNSTLKN 19  
Db 44 NNYKMHLLTTSSNSTASN 61

## RESULT 26

US-08-690-102A-4  
; Sequence 4, Application US/08690102A  
; Patent No. 5789554  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, Shui-on  
; APPLICANT: HANSEN, Hans  
; TITLE OF INVENTION: IMMUNOCONJUGATES AND HUMANIZED  
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR B-CELL LYMPHOMA AND LEUKEMIA CELLS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,102A  
; FILING DATE: 01-JUL-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,576  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAXE, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 18733/463/1MIN  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 116 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-690-102A-4

Query Match 38.4%; Score 43; DB 1; Length 116;  
Best Local Similarity 47.1%; Pred. No. 26;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYKNSSNSTL 17  
Db 54 RNDYTEYNQNFQDKATL 70

## RESULT 27

US-09-127-902-4  
; Sequence 4, Application US/09127902  
; Patent No. 6187287  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, Shui-on  
; APPLICANT: HANSEN, Hans  
; TITLE OF INVENTION: IMMUNOCONJUGATES AND HUMANIZED  
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR B-CELL LYMPHOMA AND LEUKEMIA CELLS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/127,902  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/690,102  
; FILING DATE: 01-JUL-1996  
; APPLICATION NUMBER: US 08/289,576  
; FILING DATE: 12-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAXE, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 18733/463/1MIN  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 116 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-127-902-4

Query Match 38.4%; Score 43; DB 3; Length 116;  
Best Local Similarity 47.1%; Pred. No. 26;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYKNSSNSTL 17  
Db 54 RNDYTEYNQNFQDKATL 70

## RESULT 28

US-09-155-107-4  
; Sequence 4, Application US/09155107  
; Patent No. 6254868  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, Shui-on  
; APPLICANT: HANSEN, Hans  
; APPLICANT: QU, Zhengxing  
; TITLE OF INVENTION: GLYCOSYLATED HUMANIZED B-CELL SPECIFIC ANTIBODIES

```
; FILE REFERENCE: 018733/0879
; CURRENT APPLICATION NUMBER: US/09/155,107
; CURRENT FILING DATE: 1998-11-17
; EARLIER APPLICATION NUMBER: WO PCT/US97/04196
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: US 60/013,709
; EARLIER FILING DATE: 1996-03-20
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Murine
; US-09-155-107-4

Query Match      38.4%; Score 43; DB 3; Length 116;
Best Local Similarity 47.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTL 17
   ||| ||| :||
Db 54 RNDYTEYNQPKKATL 70

RESULT 29
PCT-US95-09641-4
; Sequence 4, Application PC/TUS9509641
; GENERAL INFORMATION:
; APPLICANT: IMMUNOCONJUGATES AND HUMANIZED
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR B-CELL LYMPHOMA AND LEUKEMIA CELLS
; NUMBER OF SEQUENCES: 21
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09641
; FILING DATE: 11-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,576
; FILING DATE: 12-AUG-1994
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-09641-4

Query Match      38.4%; Score 43; DB 5; Length 116;
Best Local Similarity 47.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTL 17
   ||| ||| :||
Db 54 RNDYTEYNQPKKATL 70

RESULT 30
PCT-US96-07709-22
; Sequence 22, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
```

```
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-07709-22

Query Match      38.4%; Score 43; DB 5; Length 169;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 5 KLYNKSSNSTLKNLGE 22
   ||| ||| :||
Db 129 KLYENNPNNMTWKVAGQ 146

RESULT 31
US-09-583-110-3991
; Sequence 3991, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3991
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-583-110-3991

Query Match      38.4%; Score 43; DB 4; Length 185;
Best Local Similarity 47.4%; Pred. No. 44;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DYKLYNKSSNSTLKNLG 21
   ||| ||| :||
Db 64 DFEKYNKYILSESNNLG 82

RESULT 32
US-09-270-767-60029
; Sequence 60029, Application US/09270767
; Patent No. 6703491
```

```

; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60029
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-60029

Query Match      38.4%; Score 43; DB 4; Length 195;
Best Local Similarity 57.1%; Pred. No. 47;
Matches      8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      4 YKLYNKSSNSTL 17
Db      41 YALYKNKPKSDTL 54

RESULT 33
PCT-US96-07709-19
; Sequence 19, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-07709-19

Query Match      38.4%; Score 43; DB 5; Length 197;
Best Local Similarity 50.0%; Pred. No. 48;
Matches      9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy      5 KLYNKSSNSTLKNLGE 22
Db      157 KLYENPNNTWKVAGQ 174

RESULT 34
PCT-US96-07709-19
; Sequence 30, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-07709-19

Query Match      38.4%; Score 43; DB 5; Length 204;
Best Local Similarity 50.0%; Pred. No. 50;
Matches      9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy      5 KLYNKSSNSTLKNLGE 22
Db      113 KLYENPNNTWKVAGQ 130

RESULT 35
PCT-US96-07709-30
; Sequence 30, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:

```

```

PCT-US96-07709-33
; Sequence 33, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-07709-33

Query Match      38.4%; Score 43; DB 5; Length 204;
Best Local Similarity 50.0%; Pred. No. 50;
Matches      9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy      5 KLYNKSSNSTLKNLGE 22
Db      113 KLYENPNNTWKVAGQ 130

RESULT 35
PCT-US96-07709-30
; Sequence 30, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:

```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 220 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-07709-30

Query Match 38.4%; Score 43; DB 5; Length 220;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 5 KLYKNSSNSTLKNLGE 22
Db 129 KLYENPNNTWKVAGQ 146

RESULT 36
US-09-541-759-1
; Sequence 1, Application US/09541759
; Patent No. 6723322
; GENERAL INFORMATION:
; APPLICANT: Lustigman, Sara
; APPLICANT: Pearlman, Eric
; APPLICANT: Unnasch, Thomas
; TITLE OF INVENTION: ANGIOGENIC ONCHOCERCA VOLVULUS PROTEINS AND USES THEREOF
; FILE REFERENCE: 63475/252
; CURRENT APPLICATION NUMBER: US/09/541,759
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Onchocerca volvulus
US-09-541-759-1

Query Match 38.4%; Score 43; DB 4; Length 223;
Best Local Similarity 50.0%; Pred. No. 55;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 5 KLYKNSSNSTLKNLGE 22
Db 129 KLYENPNNTWKVAGQ 146

RESULT 37
US-09-270-767-44586
; Sequence 44586, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44586
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44586
```

```
Query Match 38.4%; Score 43; DB 4; Length 244;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 KLYKNSSNSTLKNLGE 15
Db 100 YNLYAKNNNSNS 111

RESULT 38
PCT-US96-07709-25
; Sequence 25, Application PC/TUS9607709
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
; TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07709
; FILING DATE: 23-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-30-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-07709-25
```

```
Query Match 38.4%; Score 43; DB 5; Length 248;
Best Local Similarity 50.0%; Pred. No. 63;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 5 KLYKNSSNSTLKNLGE 22
Db 157 KLYENPNNTWKVAGQ 174

RESULT 39
US-09-134-000C-4344
; Sequence 4344, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lyttu Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
```



```
; SEQ ID NO 4344
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4344

Query Match      38.4%; Score 43; DB 4; Length 281;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy      1 RSDYK--LYNKNSSSN 14
      |||:| ||| | :|
Db      246 RSDWKDYLYNVNKSNN 261

RESULT 40
US-09-270-767-44579
; Sequence 44579, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44579
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44579

Query Match      38.4%; Score 43; DB 4; Length 307;
Best Local Similarity 57.1%; Pred. No. 80;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      4 YKLYNKNSSSNSTL 17
      ||||| : ||
Db      41 YALYNKNPKSDTL 54

Search completed: November 24, 2004, 09:31:39
Job time : 21.3696 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:07:49 ; Search time 95.8913 Seconds  
(without alignments)  
82.302 Million cell updates/sec

Title: US-09-719-379A-2  
Perfect score: 112  
Sequence: 1 RSDYKLYKNSSNSTLKNLGE 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_23Sep04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 112   | 100.0       | 22     | 3     | AAY79988 Non-typea |
| 2          | 112   | 100.0       | 22     | 3     | AAY79983 Non-typea |
| 3          | 112   | 100.0       | 22     | 3     | AAY79974 Non-typea |
| 4          | 112   | 100.0       | 22     | 4     | AAB47440 LBI(f) co |
| 5          | 112   | 100.0       | 31     | 4     | AAB47444 Entire 3r |
| 6          | 112   | 100.0       | 464    | 3     | AAY79993 Plasmid L |
| 7          | 97.5  | 87.1        | 21     | 3     | AAY79977 Non-typea |
| 8          | 97.5  | 87.1        | 21     | 3     | AAY79975 Non-typea |
| 9          | 86.5  | 77.2        | 21     | 3     | AAY79976 Non-typea |
| 10         | 83    | 74.1        | 20     | 3     | AAY79979 Non-typea |
| 11         | 81.5  | 72.8        | 19     | 3     | AAY79978 Non-typea |
| 12         | 81.5  | 72.8        | 19     | 3     | AAY79980 Non-typea |
| 13         | 81.5  | 72.8        | 19     | 3     | AAY79984 Non-typea |
| 14         | 81.5  | 72.8        | 19     | 4     | AAB47442 LBI(f) co |
| 15         | 81.5  | 72.8        | 28     | 4     | AAB47446 Entire 3r |
| 16         | 81.5  | 72.8        | 353    | 5     | ABG80420 Haemophil |
| 17         | 76.5  | 68.3        | 353    | 4     | AAB47447 MOMP p5.  |
| 18         | 76    | 67.9        | 20     | 3     | AAY79980 Non-typea |
| 19         | 48    | 42.9        | 706    | 6     | ABU50194 Protein e |
| 20         | 48    | 42.9        | 1390   | 8     | ADN46970 Thermoco  |
| 21         | 48    | 42.9        | 1878   | 4     | ABB61852 Drosophil |
| 22         | 47    | 42.0        | 308    | 7     | ADD71579 Human uri |
| 23         | 47    | 42.0        | 510    | 2     | RAY05901 Tobacco f |
| 24         | 46    | 41.1        | 324    | 3     | AKG31785 Arabidops |
| 25         | 46    | 41.1        | 324    | 6     | ABP81228 Arabidops |

|    |      |      |      |   |                    |
|----|------|------|------|---|--------------------|
| 26 | 46   | 41.1 | 332  | 3 | AAB18299 Plasmodiu |
| 27 | 46   | 41.1 | 2052 | 6 | ABU23678 Protein e |
| 28 | 45   | 40.2 | 60   | 7 | ADF04538 Bacterial |
| 29 | 45   | 40.2 | 225  | 4 | AAB62090 T. indica |
| 30 | 45   | 40.2 | 225  | 6 | AAE30847 Tilletia  |
| 31 | 45   | 40.2 | 338  | 5 | ABG93043 S. cerevi |
| 32 | 45   | 40.2 | 423  | 5 | ABP28260 Streptoco |
| 33 | 45   | 40.2 | 509  | 2 | AAW20085 Helicobac |
| 34 | 45   | 40.2 | 593  | 2 | AAW20918 H. pylori |
| 35 | 45   | 40.2 | 657  | 4 | ABG14461 Novel hum |
| 36 | 45   | 40.2 | 733  | 6 | ADA34312 Acinetoba |
| 37 | 45   | 40.2 | 784  | 4 | ABB71161 Drosophil |
| 38 | 45   | 40.2 | 996  | 4 | AB84950 Novel hum  |
| 39 | 45   | 40.2 | 1049 | 5 | ABG69624 Human sec |
| 40 | 45   | 40.2 | 1758 | 6 | ABRS2794 Protein s |
| 41 | 45   | 40.2 | 1758 | 7 | ADKG2164 Disease t |
| 42 | 45   | 40.2 | 1796 | 5 | ABG93090 S. cerevi |
| 43 | 45   | 40.2 | 1841 | 5 | ADH32883 Yeast smO |
| 44 | 44.5 | 39.7 | 1237 | 2 | AAW44729 Chicken p |
| 45 | 44.5 | 39.7 | 1237 | 2 | AAW89347 Chicken t |

ALIGNMENTS

RESULT 1  
AAY79988

ID AAY79988 standard; peptide; 22 AA.

XX AAY79988;

XX 15-MAY-2000 (first entry)

XX Non-typeable H. influenzae 1715MEE Group 2a type peptide.

XX Vaccine; non-typeable Haemophilus influenzae; ntl; infection;  
chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;  
conjunctivitis; lower respiratory tract infection.

OS Haemophilus influenzae.

XX W09964067-A2.

XX 16-DEC-1999.

XX 28-MAY-1999; 99WO-US011980.

XX 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
(OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
Haemophilus influenza.

XX Claim 3; Page 46; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
peptides (LBI(f) peptides) of P5-like fimbria proteins from various  
Haemophilus influenza strains. The peptides are used for diagnosis,  
prevention, and treatment of Haemophilus influenza infections, such as  
otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
infection. The peptides may also be used in vaccines against H.  
influenzae. Antibodies and probes from the present invention can be used  
for diagnosis of H. influenza infection. AAY79955 to AAY79993, and  
AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
the present invention

```

SQ Sequence 22 AA;
Query Match      100.0%; Score 112; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLGE 22
Db 1 RSDYKLYNKNSSNSTLKNLGE 22

RESULT 2
AAAY79983
ID AAY79983 standard; peptide; 22 AA.
XX
AC AAY79983;
XX
DT 15-MAY-2000 (first entry)
XX
DE Non-typeable H. influenzae group 2a LB1(f) peptide N1715MEE.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX
PN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
OS WPI; 2000-116457/10.
XX
PN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
PT Haemophilus influenza.
XX
PS Example 1; Page 30; 68pp; English.
XX
CC The present invention describes antigenic P5-like fimbria subunit
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
CC Haemophilus influenzae strains. The peptides are used for diagnosis,
CC prevention, and treatment of Haemophilus influenzae infections, such as
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
CC infection. The peptides may also be used in vaccines against H.
CC influenzae. Antibodies and probes from the present invention can be used
CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993; and
CC AA291201 to AA291252, represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 22 AA;
Query Match      100.0%; Score 112; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLGE 22
Db 1 RSDYKLYNKNSSNSTLKNLGE 22

RESULT 4
AAB47440
ID AAB47440 standard; peptide; 22 AA.
XX
AC AAB47440;
XX
DT 31-OCT-2001 (first entry)
XX
DE LB1(f) containing peptide from strain nH1-1715MEE (Group 2a type).
XX
KW surface exposed loop; major outer membrane protein P5; MOMP P5;
KW non-typeable H. influenzae; nH1; LB1(f) peptide; B cell epitope;
KW otitis media; sinusitis; conjunctivitis;
KW lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX

```

```

PN WO200161013-A1.
PD 23-AUG-2001.
XX
XX 13-FEB-2001; 2001WO-EP001556.
XX
XX 15-FEB-2000; 2000GB-00003502.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Denoel P, Poolman J, Thonnard J;
XX WPI; 2001-522599/57.
XX
XX Recombinant bacterial outer membrane protein where one or more surface-
XX exposed loops are modified is useful as a vaccine to prevent or treat
XX Haemophilus influenzae infection or associated disease, e.g., otitis
XX media and conjunctivitis.
XX
XX Claim 1; Page 26; 29pp; English.
XX
XX The sequences given in AAB47439-46 represent peptides which may be used
XX to replace one or more surface exposed loops of major outer membrane
XX protein P5 (MOMP P5) of non-typeable H. influenzae (ntHi). Each of these
XX peptides contain an LBI(f) peptide which is a 19 amino acid peptide
XX derived from the sequence of MOMP P5 from strain nHil128, representing
XX amino acids Arg117 to Gly135. This peptide represents the third exposed
XX loop of P5 and is a potential B cell epitope. The loops of the invention
XX are modified in terms of being in a non-native environment in the
XX recombinant outer membrane protein. The modified MOMP P5 may be used to
XX induce an immune response in a mammal to prevent or treat Haemophilus
XX influenzae infection or associated disease, e.g., otitis media,
XX sinusitis, conjunctivitis, or lower respiratory tract infection
XX
XX Sequence 22 AA;
XX
XX Query Match 100.0%; Score 112; DB 4; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-10;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RSDYKLYNKSSNSTLKNLGE 22
XX |||||||||||||||||||
XX DB 1 RSDYKLYNKSSNSTLKNLGE 22
XX
XX RESULT 6
XX AAY79993
XX ID AAY79993 standard; protein; 464 AA.
XX
XX AC AAY79993;
XX
XX DT 15-MAY-2000 (first entry)
XX
XX DE Plasmid LPD-LBI-III protein sequence.
XX
XX KW Vaccine; non-typeable Haemophilus influenzae; ntHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX
XX OS Haemophilus influenzae.
XX
XX OS Synthetic.
XX
XX PN WO9964067-A2.
XX
XX PD 16-DEC-1999.
XX
XX PF 28-MAY-1999; 99WO-US011980.
XX
XX PR 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX N-PSDB; AAZ91252.
XX
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.
XX
XX Claim 14; Fig 5; 68pp; English.
XX

```

CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention

XX Sequence 464 AA;

Query Match 100.0%; Score 112; DB 3; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 7.5e-09;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22  
 |||||  
 Db 377 RSDYKLYNKSSNSTLKNLGE 398

#### RESULT 7

AA79977  
 ID AAY79977 standard; peptide; 21 AA.

XX AC AAY79977;

XX DT 15-MAY-2000 (first entry)

XX DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-506.

XX KW Vaccine; non-typeable Haemophilus influenzae; nH; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX XW Haemophilus influenzae.

XX OS WO9964067-A2.

XX PN 16-DEC-1999.

XX PD 28-MAY-1999; 99WO-US011980.

XX PF 11-JUN-1998; 98GB-00012613.

XX PR (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.  
 (OHIS) UNIV OHIO STATE RES FOUND.

XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX PT WPI; 2000-116457/10.

XX DR Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenza.

XX PS Example 1; Page 30; 68pp; English.

XX CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention

XX SQ Sequence 21 AA;

Query Match 87.1%; Score 97.5; DB 3; Length 21;

Best Local Similarity 95.5%; Pred. No. 3.5e-08;  
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22  
 |||||  
 Db 1 RSDYKLYNK-SSNSTLKNLGE 21

#### RESULT 8

AA79975  
 ID AAY79975 standard; peptide; 21 AA.

XX AC AAY79975;

XX DT 15-MAY-2000 (first entry)

XX DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-492.

XX KW Vaccine; non-typeable Haemophilus influenzae; nH; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX OS Haemophilus influenzae.

XX PN WO9964067-A2.

XX PD 16-DEC-1999.

XX PF 28-MAY-1999; 99WO-US011980.

XX PR 11-JUN-1998; 98GB-00012613.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.  
 (OHIS) UNIV OHIO STATE RES FOUND.

XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX DR WPI; 2000-116457/10.

XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenza.

XX PS Example 1; Page 30; 68pp; English.

XX CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention

XX SQ Sequence 21 AA;

Query Match 87.1%; Score 97.5; DB 3; Length 21;  
 Best Local Similarity 95.5%; Pred. No. 3.5e-08;  
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22  
 |||||  
 Db 1 RSDYKLYNK-SSNSTLKNLGE 21

#### RESULT 9

AA79976  
 ID AAY79976 standard; peptide; 21 AA.

XX AC AAY79976;

XX CC

```

DT 15-MAY-2000 (first entry)
DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-502.
XX Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX Haemophilus influenzae.
XX OS WO9964067-A2.
XX PN 16-DEC-1999.
XX 28-MAY-1999; 99WO-US011980.
XX 11-JUN-1998; 98GB-00012613.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.
XX Example 1; Page 30; 68pp; English.
XX The present invention describes antigenic P5-like fimbria subunit
XX peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX Haemophilus influenzae strains. The peptides are used for diagnosis,
XX prevention, and treatment of Haemophilus influenzae infections, such
XX as otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX infection. The peptides may also be used in vaccines against H.
XX influenzae. Antibodies and probes from the present invention can be used
XX for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX AA291201 to AA291252, represent sequences used in the exemplification of
XX the present invention
XX SQ Sequence 21 AA;
XX Query Match 77.2%; Score 86.5; DB 3; Length 21;
XX Best Local Similarity 86.4%; Pred. No. 1.9e-06;
XX Matches 19; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
XX Qy 1 RSDYKLYNKSSNSTLKNLGE 22
XX Db 1 RSDYKLYNKSSN-TLKLGE 21
XX RESULT 10
XX AAY79979
XX ID AAY79979 standard; peptide; 20 AA.
XX AC AAY79979;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 2 LB1(f) peptide N165NP.
XX Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX Haemophilus influenzae.
XX OS WO9964067-A2.
XX PN 16-DEC-1999.
XX 28-MAY-1999; 99WO-US011980.
XX 11-JUN-1998; 98GB-00012613.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.
XX Example 1; Page 30; 68pp; English.
XX The present invention describes antigenic P5-like fimbria subunit
XX peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX Haemophilus influenzae strains. The peptides are used for diagnosis,
XX prevention, and treatment of Haemophilus influenzae infections, such
XX as otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX infection. The peptides may also be used in vaccines against H.
XX influenzae. Antibodies and probes from the present invention can be used
XX for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX AA291201 to AA291252, represent sequences used in the exemplification of
XX the present invention
XX SQ Sequence 21 AA;
XX Query Match 74.1%; Score 83; DB 3; Length 20;
XX Best Local Similarity 81.8%; Pred. No. 6.4e-06;
XX Matches 18; Conservative 2; Mismatches 0; Indels 2; Gaps 1;
XX Qy 1 RSDYKLYNKSSNSTLKNLGE 22
XX Db 1 RSDYKLYNKSS--NTLKLGE 20
XX RESULT 11
XX AAY79978
XX ID AAY79978 standard; peptide; 19 AA.
XX AC AAY79978;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 2 LB1(f) peptide N1236NEE.
XX Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX Haemophilus influenzae.
XX OS WO9964067-A2.
XX PN 16-DEC-1999.
XX 28-MAY-1999; 99WO-US011980.
XX 11-JUN-1998; 98GB-00012613.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.

```

```

XX PS Example 1; Page 30; 68pp; English.
XX CC The present invention describes antigenic P5-like fimbrin subunit
XX CC peptides (Lb1(f) peptides) of P5-like fimbrin proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention
XX SQ Sequence 19 AA;

    Query Match      72.8%; Score 81.5; DB 3; Length 19;
    Best Local Similarity 81.8%; Pred. No. 1e-05;
    Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 1 RSDYKLYNKNS---STLKDLE 19

RESULT 12
AAY79990
ID AAY79990 standard; peptide; 19 AA.
XX AC AAY79990;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae 183NP Group 2b type peptide.
XX KW Vaccine; non-typeable Haemophilus influenzae; ntHi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
XX KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.
XX PF 28-MAY-1999; 99WO-US011980.
XX PR 11-JUN-1998; 98GB-00012613.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.
XX PF 28-MAY-1999; 99WO-US011980.
XX PR 11-JUN-1998; 98GB-00012613.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX DR WPI; 2000-116457/10.
XX PT Novel antigenic P5-like fimbrin subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PS Claim 5; Page 46; 69pp; English.
XX CC The present invention describes antigenic P5-like fimbrin subunit
XX CC peptides (Lb1(f) peptides) of P5-like fimbrin proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention

```

```

SQ Sequence 19 AA;

    Query Match      72.8%; Score 81.5; DB 3; Length 19;
    Best Local Similarity 81.8%; Pred. No. 1e-05;
    Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 1 RSDYKLYNKNS---STLKDLE 19

RESULT 13
AAY79984
ID AAY79984 standard; peptide; 19 AA.
XX AC AAY79984;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 2b Lb1(f) peptide NTHI-183.
XX KW Vaccine; non-typeable Haemophilus influenzae; ntHi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
XX KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.
XX PF 28-MAY-1999; 99WO-US011980.
XX PR 11-JUN-1998; 98GB-00012613.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX DR WPI; 2000-116457/10.
XX PT Novel antigenic P5-like fimbrin subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PS Example 1; Page 30; 68pp; English.
XX CC The present invention describes antigenic P5-like fimbrin subunit
XX CC peptides (Lb1(f) peptides) of P5-like fimbrin proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX CC the present invention
XX SQ Sequence 19 AA;

    Query Match      72.8%; Score 81.5; DB 3; Length 19;
    Best Local Similarity 81.8%; Pred. No. 1e-05;
    Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKSSNSTLKNLGE 22
Db 1 RSDYKLYNKNS---STLKDLE 19

RESULT 14
AAB47442
ID AAB47442 standard; peptide; 19 AA.

```



```

XX AAB47442;
AC
XX 31-OCT-2001 (first entry)
DT
XX LBI(f) cont. peptide from strain nH1-183NP (Group 2b type).
DE
XX surface exposed loop; major outer membrane protein P5; MOMP P5;
KW non-typeable H. influenzae; nH1; LBI(f) peptide; B cell epitope;
KW otitis media; sinusitis; conjunctivitis;
KW lower respiratory tract infection.
XX Haemophilus influenzae.
OS
XX WO200161013-A1.
XX
XX 23-AUG-2001.
PD
XX
XX 13-FEB-2001; 2001WO-EP001556.
PF
XX
XX 15-FEB-2000; 2000GB-00003502.
PR
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Denoel P, Poolman J, Thonnard J;
PI
XX WPI; 2001-522599/57.
XX
XX Recombinant bacterial outer membrane protein where one or more surface-
PT exposed loops are modified is useful as a vaccine to prevent or treat
PT Haemophilus influenzae infection or associated disease, e.g., otitis
PT media and conjunctivitis.
XX
XX Claim 1; Page 26; 29pp; English.
XX
XX The sequences given in AAB47439-46 represent peptides which may be used
CC to replace one or more surface exposed loops of major outer membrane
CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these
CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide
CC derived from the sequence of MOMP P5 from strain nH1128, representing
CC amino acids Arg117 to Gly135. This peptide represents the third exposed
CC loop of P5 and is a potential B cell epitope. The loops of the invention
CC are modified in terms of being in a non-native environment in the
CC recombinant outer membrane protein. The modified MOMP P5 may be used to
CC induce an immune response in a mammal to prevent or treat Haemophilus
CC influenzae infection or associated disease, e.g., otitis media,
CC sinusitis, conjunctivitis, or lower respiratory tract infection
XX
XX Sequence 19 AA;
SQ
Query Match 72.8%; Score 81.5; DB 4; Length 19;
Best Local Similarity 81.8%; Pred. No. 1e-05;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNSNSTLKNLGE 22
Db ||||| ||||| |||||
1 RSDYKLYNKNS---STLKD LGE 19

RESULT 15
AAB47446
XX AAB47446 standard; peptide; 28 AA.
XX
XX AAB47446;
AC
XX
XX 31-OCT-2001 (first entry)
DT
XX Entire 3rd loop from strain nH1-183NP (Group 2b type).
DE
XX surface exposed loop; major outer membrane protein P5; MOMP P5;
KW non-typeable H. influenzae; nH1; LBI(f) peptide; B cell epitope;
KW otitis media; sinusitis; conjunctivitis;
KW lower respiratory tract infection.

```

```

XX Haemophilus influenzae.
OS
XX WO200161013-A1.
XX
XX 23-AUG-2001.
PD
XX
XX 13-FEB-2001; 2001WO-EP001556.
PF
XX
XX 15-FEB-2000; 2000GB-00003502.
PR
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Denoel P, Poolman J, Thonnard J;
PI
XX WPI; 2001-522599/57.
XX
XX Recombinant bacterial outer membrane protein where one or more surface-
PT exposed loops are modified is useful as a vaccine to prevent or treat
PT Haemophilus influenzae infection or associated disease, e.g., otitis
PT media and conjunctivitis.
XX
XX Claim 2; Page 26; 29pp; English.
XX
XX The sequences given in AAB47439-46 represent peptides which may be used
CC to replace one or more surface exposed loops of major outer membrane
CC protein P5 (MOMP P5) of non-typeable H. influenzae (nH1). Each of these
CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide
CC derived from the sequence of MOMP P5 from strain nH1128, representing
CC amino acids Arg117 to Gly135. This peptide represents the third exposed
CC loop of P5 and is a potential B cell epitope. The loops of the invention
CC are modified in terms of being in a non-native environment in the
CC recombinant outer membrane protein. The modified MOMP P5 may be used to
CC induce an immune response in a mammal to prevent or treat Haemophilus
CC influenzae infection or associated disease, e.g., otitis media,
CC sinusitis, conjunctivitis, or lower respiratory tract infection
XX
XX Sequence 28 AA;
SQ
Query Match 72.8%; Score 81.5; DB 4; Length 28;
Best Local Similarity 81.8%; Pred. No. 1.6e-05;
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNSNSTLKNLGE 22
Db ||||| ||||| |||||
3 RSDYKLYNKNS---STLKD LGE 21

RESULT 16
ABG80420
ID ABG80420 standard; protein; 353 AA.
XX
XX AC ABG80420;
XX
XX 29-NOV-2002 (first entry)
DT
XX Haemophilus influenzae P5 protein.
DE
XX
XX Hyperblebbing; Gram-negative bacterium; genetically modified; Tol gene;
KW peptidoglycan-binding; peptidoglycan-associated site; outer-membrane;
KW bacterial infection; vesicle-shedding; B1eb; filter sterilised;
KW detergent; deoxycholate; homogeneity; antibacterial; vaccine; H1RD; P5;
KW P6; PCP.
XX
XX Haemophilus influenzae.
OS
XX
XX WO200262378-A2.
XX
XX 15-AUG-2002.
PD
XX
XX 08-FEB-2002; 2002WO-EP001361.
PF
XX
XX 08-FEB-2001; 2001GB-00003171.
PR

```

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Berthet FJ, Denoel P, Neyt CA, Poolman J, Thonnard J;

XX XX

XX DR WPI; 2002-657509/70.

XX DR N-PSDB; ABS66193.

XX PT Novel hyperblebbing Gram-negative bacterium that has been genetically

XX PT modified by down-regulating expression of Tol genes, and/or attenuating

XX PT peptidoglycan-binding activity useful for treating bacterial infection.

XX PS Disclosure; Page 58; 71pp; English.

XX XX

CC The present invention relates to a new hyperblebbing Gram-negative

CC bacterium genetically modified by one or more processes selected from

CC down-regulating expression of one or more Tol genes and attenuating the

CC peptidoglycan-binding activity by mutation of one or more gene(s)

CC encoding a protein comprising a peptidoglycan-associated site. The

CC invention is useful in a method of treatment of the human or animal body.

CC The invention is also useful for protecting an individual against a

CC bacterial infection. The invention has improved outer-membrane vesicle-

CC shedding properties. Blebs are more easily made in higher yield from the

CC invention, and are more homogeneous in nature and can be more readily

CC filter sterilised. The blebs can be made and harvested without the use of

CC detergents such as deoxycholate, thus obviating chromatography

CC purification and ultra centrifugation steps. Vesicles prepared from the

CC invention have reduced particle size (allowing sterile filtration through

CC 0.22 µm pores), increased batch homogeneity, and a superior yield. The

CC present amino acid sequence represents a Haemophilus influenzae protein,

CC as described in the invention

XX XX

SQ Sequence 353 AA;

Query Match 72.8%; Score 81.5; DB 5; Length 353;

Best Local Similarity 81.8%; Pred. No. 0.00035;

Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNSNSTLKNLGE 22

Db 132 RSDYKLYNKNS--STLKLGE 150

RESULT 17

AAB47447

ID AAB47447 standard; protein; 353 AA.

XX AC AAB47447;

XX DT 31-OCT-2001 (first entry)

XX DE MOMP P5.

XX XX

KW surface exposed loop; major outer membrane protein P5; MOMP P5;

KW non-typeable H. influenzae; nH1; LB1(f) peptide; B cell epitope;

KW otitis media; sinusitis; conjunctivitis;

KW lower respiratory tract infection.

XX OS Haemophilus influenzae.

XX FH Key Location/Qualifiers

FT Domain 38..57

FT /label= Loop 1

FT /note= "Extracellular domain"

FT Domain 89..100

FT /label= Loop 2

FT /note= "Extracellular domain"

FT Domain 136..150

FT /label= Loop 3

FT /note= "Extracellular domain"

FT Domain 181..204

FT /label= Loop 4

FT /note= "Extracellular domain"

XX WO200161013-A1.

XX PD 23-AUG-2001.

XX PF 13-FEB-2001; 2001WO-EP001556.

XX PR 15-FEB-2000; 2000GB-00003502.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Berthet FJ, Denoel P, Poolman J, Thonnard J;

XX DR WPI; 2001-522599/57.

XX XX

PT Recombinant bacterial outer membrane protein where one or more surface-

PT exposed loops are modified is useful as a vaccine to prevent or treat

PT Haemophilus influenzae infection or associated disease, e.g., otitis

XX media and conjunctivitis.

XX PS Disclosure; Fig 1; 29pp; English.

XX CC This sequence represents the major outer membrane protein P5 of non-

CC typeable H. influenzae. One or more surface exposed loops of this protein

CC may be replaced with a modified peptide of the invention. Each of these

CC peptides contain an LB1(f) peptide which is a 19 amino acid peptide

CC derived from the sequence of MOMP P5 from strain nH1128, representing

CC amino acids Arg117 to Gly135. This peptide represents the third exposed

CC loop of P5 and is a potential B cell epitope. The loops of the invention

CC are modified in terms of being in a non-native environment in the

CC recombinant outer membrane protein. The modified MOMP P5 may be used to

CC induce an immune response in a mammal to prevent or treat Haemophilus

CC influenzae infection or associated disease, e.g., otitis media,

CC sinusitis, conjunctivitis, or lower respiratory tract infection

XX XX

SQ Sequence 353 AA;

Query Match 68.3%; Score 76.5; DB 4; Length 353;

Best Local Similarity 77.3%; Pred. No. 0.0021;

Matches 17; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 1 RSDYKLYNKNSNSTLKNLGE 22

Db 132 RSDYKLYNKNS--STLKLGE 150

RESULT 18

AAY79980

ID AAY79980 standard; peptide; 20 AA.

XX AC AAY79980;

XX DT 15-MAY-2000 (first entry)

XX XX

DE Non-typeable H. influenzae group 2 LB1(f) peptide NTHI-495.

XX XX

KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;

KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;

KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;

KW conjunctivitis; lower respiratory tract infection.

XX OS Haemophilus influenzae.

XX FH Key Location/Qualifiers

FT Domain 38..57

FT /label= Loop 1

FT /note= "Extracellular domain"

FT Domain 89..100

FT /label= Loop 2

FT /note= "Extracellular domain"

FT Domain 136..150

FT /label= Loop 3

FT /note= "Extracellular domain"

FT Domain 181..204

FT /label= Loop 4

FT /note= "Extracellular domain"

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX (OHIS ) UNIV OHIO STATE RES FOUND.

XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
XX DR WPI; 2000-116457/10.  
XX PT Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
XX PT Haemophilus influenza.  
XX Example 1; Page 30; 68pp; English.  
XX The present invention describes antigenic P5-like fimbrin subunit  
XX peptides (Lb1(f) peptides) of P5-like fimbrin proteins from various  
XX Haemophilus influenzae strains. The peptides are used for diagnosis,  
XX prevention, and treatment of Haemophilus influenzae infections, such as  
XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
XX infection. The peptides may also be used in vaccines against H.  
XX influenzae. Antibodies and probes from the present invention can be used  
XX for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and  
XX AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
XX the present invention  
XX SQ Sequence 20 AA;  
Query Match 67.9%; Score 76; DB 3; Length 20;  
Best Local Similarity 77.3%; Pred. No. 8.1e-05;  
Matches 17; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 1 RSDYKLYNKSSNSTLKNLGE 22  
DB 1 RSDYKLYNKSS--DALKKLGE 20  
RESULT 19  
ABU50194  
ID ABU50194 standard; protein; 706 AA.  
XX AC ABU50194;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #35721.  
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Yersinia pestis.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 23-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA54064.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 78118; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 706 AA;  
Query Match 42.9%; Score 48; DB 6; Length 706;  
Best Local Similarity 40.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
QY 3 DYKLYNKSSNSTLKNLGE 22  
DB 543 DYERFNVNKASNTVLSSLNQ 562  
RESULT 20  
ADN46970  
ID ADN46970 standard; protein; 1390 AA.  
XX AC ADN46970;  
XX DT 01-JUL-2004 (first entry)  
XX DE Thermococcus kodakaraensis KOD1 protein sequence SeqID848.  
XX gene disruption; gene targeting; marker gene; transformation;  
XX homologous recombination; hyperthermostable archaeobacterium; KOD1;  
XX gene structure; gene function; enzyme activity; medicine;  
XX forensic science; food; drug inspection; molecular biology; immunology.  
XX OS Thermococcus kodakaraensis.  
XX WO2004022735-A1.  
XX 18-MAR-2004.  
XX 29-AUG-2003; 2003WO-IB003597.  
XX 30-AUG-2002; 2002JP-00319011.  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX Imanaka T, Atomi H;  
XX WPI; 2004-257583/24.  
XX

PT Method for disrupting targeted gene in genome of organism particularly  
PT thermostable bacterium and with genome chips for analysis, applicable in  
PT studying gene structure and functions.  
XX  
PS Claim 9; SEQ ID NO 848; 598pp; Japanese.  
XX  
CC This invention relates to a novel method for targeting disruption of an  
CC arbitrary gene in a genome of an organism which comprises providing the  
CC whole sequential data of the genome of such organism, selecting at least  
CC 1 arbitrary region in the sequence, providing a vector that contains a  
CC sequence homologous with the selected region and a marker gene.  
CC transformation, and homologous recombination. The genome is preferably  
CC the genome of a hyperthermostable archaeobacterium, particularly  
CC Thermococcus kodakarensis KOD1. The method is for targeting the  
CC disruption of a gene in the genome of an organism, which is applicable in  
CC studying gene structure and functions as well as enzyme activities of  
CC encoded proteins and useful in medicine, forensic science, food or drug  
CC inspection, molecular biology and immunology. With this method, the  
CC disruption of a gene at an arbitrary position in a genome can be achieved  
CC efficiently and reliably. The present sequence is that of a protein  
CC encoded by the genome of Thermococcus kodakarensis which was derived  
CC using the method of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1390 AA;  
  
Query Match 42.9%; Score 48; DB 8; Length 1390;  
Best Local Similarity 38.5%; Pred. No. 3.4e+02;  
Matches 10; Conservative 7; Mismatches 5; Indels 4; Gaps 1;  
  
Qy 1 RSDYK---LYNKNSSNSTLKNLGE 22  
||||: |||: ||: |||:  
736 RSDVEGMEYNEGYNSDVVPNLGK 761  
  
RESULT 21  
ABB61852  
ID ABB61852 standard; protein; 1878 AA.  
XX  
AC ABB61852;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 12348.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PS 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-ESDS; ABL05955.  
XX  
CC New isolated nucleic acid detection reagent for detecting 1000 or more  
CC genes from Drosophila and for elucidating cell signalling and cell-cell  
CC interactions.  
XX  
PS Disclosure; SEQ ID NO 12348; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1878 AA;  
  
Query Match 42.9%; Score 48; DB 4; Length 1878;  
Best Local Similarity 45.0%; Pred. No. 4.9e+02;  
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 1 RSDYKLYNKNSSNSTLKNL 20  
|:|:|:|:|:|:|:|:|:|:  
Db 1256 REEYLLYNTENSTTQTNI 1275  
  
RESULT 22  
ADD71579  
ID ADD71579 standard; protein; 308 AA.  
XX  
AC ADD71579;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human urinary specific protein sequence SEQ ID NO:420.  
XX  
KW human; urinary specific nucleic acid; USNA; cytostatic; gene therapy;  
KW vaccine; urinary cancer; urinary specific protein; USP; cancer;  
KW squamous cell carcinoma.  
XX  
OS Homo sapiens.  
XX  
PN WO2003060146-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 19-DEC-2002; 2002WO-US041027.  
XX  
PR 21-DEC-2001; 2001US-0342976P.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Sun Y, Liu C;  
XX  
DR WPI; 2003-618182/58.  
XX  
PT New urinary specific genes and proteins, useful in gene therapy or as  
PT vaccines for treating urinary cancer or non-cancerous urinary diseases,  
PT as well as for diagnosing, monitoring or staging such diseases.  
XX  
PS Claim 1; SEQ ID NO 420; 723pp; English.  
XX  
CC The invention relates to a novel urinary specific nucleic acid molecule  
CC (USNA). A protein encoded by a USNA of the invention has cytostatic  
CC activity. The USNA may have a use in gene therapy, and as a vaccine.  
CC These nucleic acids and polypeptides are also useful for diagnosing and  
CC monitoring the presence and metastases of urinary cancer in a patient.  
CC The antibody that specifically binds to the lung specific polypeptide is  
CC useful for determining the presence of a urinary specific protein in a  
CC sample, as well as for treating a patient with cancer, particularly by  
CC inducing an immune response against the urinary cancer cell expressing  
CC the urinary specific nucleic acid molecule or polypeptide. Particularly,  
CC these urinary specific genes and proteins are useful for identifying,  
CC diagnosing, monitoring, staging, imaging and treating urinary cancer  
CC (e.g. squamous cell carcinoma) and non-cancerous disease states in the  
CC urinary. The sequences shown in ADD71534-ADD71716 represent urinary

CC specific nucleic acids of the invention.

XX Sequence 308 AA;

Query Match 42.0%; Score 47; DB 7; Length 308;

Best Local Similarity 60.0%; Pred. No. 80;

Matches 12; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 4 YKLY--NNKSSNSTLKNLG 21

DB 33 YKLVNENRNLSSNSAKKELG 52

RESULT 23

AAAY05901

ID AAY05901 standard; protein; 510 AA.

XX AC AAY05901;

DT 02-AUG-1999 (first entry)

XX Tobacco fatty acid hydroxylase CYP94A6.

XX Fatty acid omega-hydroxylase; cytochrome P450; transgenic plant; lipid;  
KW hydroxylation; epoxidation; oilseed; vegetable oil; crop protection;  
KW omega-hydroxy acid; CYP94A6; tobacco.

XX Nicotiana tabacum.

XX Key Location/Qualifiers

FT Domain 317..328

FT /note= "haem-binding domain, corresponds to signature  
FT motif claimed in Claim 2"

XX W09918224-A1.

XX 15-APR-1999.

XX 06-OCT-1998; 98WO-IB001716.

XX 06-OCT-1997; 97US-0060960P.

XX (CNRS ) CENT NAT RECH SCI.

XX Tijet N, Pinot F, Benveniste I, Le Bouquin R, Helvig C, Batard Y;

PI Cabello-Huatao F, Werck-Reichhart D, Salaun J, Durst F;

XX WPI; 1999-264030/22.

DR N-PSDB; AAX58405.

XX Nucleic acid encoding plant fatty acid hydroxylases.

XX Example 7; Fig 26A-B; 157pp; English.

XX This sequence represents a cytochrome P450 protein, CYP94A6, of tobacco  
CC that is expected to have fatty acid hydroxylase activity since is  
CC displays the characteristic signature motif for this class of enzymes.  
CC The invention provides isolated nucleic acids (see AAX58400-06) encoding  
CC plant fatty acid hydroxylases (see AAY05896-902). Also claimed are host  
CC cells, transgenic plants and compositions consisting of the plant fatty  
CC acid hydroxylase, a process for isolating additional fatty acid  
CC hydroxylase genes from a plant, and a process of altering fatty acid  
CC composition in a plant by expressing the plant fatty acid hydroxylase in  
CC a transgenic plant, and hydroxylating or epoxidating a fatty acid  
CC substrate in the plant. Manipulating the hydroxylated fatty acid content  
CC of plants will modify resistance to drought and attack by insects and  
CC other pests. The transgenic plants may also be used as sources of  
CC hydroxylated and epoxidized fatty acids useful in the manufacture of e.g.  
CC lubricants, anti-slip agents, plasticisers, coating agents, detergents  
CC and surfactants

XX Sequence 510 AA;

SQ

Query Match 42.0%; Score 47; DB 2; Length 510;

Best Local Similarity 36.8%; Pred. No. 1.5e+02;

Matches 7; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 RSDYKLYNNKSSNSTLKN 19

DB 103 KTNFQVYQKGNHNTILKD 121

RESULT 24

AAG31785

ID AAG31785 standard; protein; 324 AA.

XX AC AAG31785;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 38230.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EF10333405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.

XX 21-APR-1999; 99US-0130077P.

XX 23-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130891P.

XX 30-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 04-MAY-1999; 99US-0132407P.

XX 05-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132485P.

XX 07-MAY-1999; 99US-0132487P.

XX 11-MAY-1999; 99US-0132863P.

XX 14-MAY-1999; 99US-0134256P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 18-MAY-1999; 99US-0134370P.

XX 19-MAY-1999; 99US-0134768P.

XX 20-MAY-1999; 99US-0134941P.

XX 21-MAY-1999; 99US-0135124P.

XX 25-MAY-1999; 99US-0135353P.

XX 27-MAY-1999; 99US-0136021P.

XX 28-MAY-1999; 99US-0136392P.

XX 01-JUN-1999; 99US-0136782P.

XX 04-JUN-1999; 99US-0137222P.

XX 07-JUN-1999; 99US-0137528P.

XX 08-JUN-1999; 99US-0137724P.

XX 10-JUN-1999; 99US-0138094P.

XX 10-JUN-1999; 99US-0138540P.

XX 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140981P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142115P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144332P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145182P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 27-JUL-1999; 99US-0145951P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.

PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 13-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160800P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 41.1%; Score 46; DB 3; Length 324;  
Best Local Similarity 38.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNIG 21  
Db 14 RESRFLFDKNGDGSITKKEIG 34

RESULT 25  
ABP81228  
ID ABP81228 standard; protein; 324 AA.  
XX AC ABP81228;  
XX AC ABP81228;  
XX DT 27-FEB-2003 (first entry)  
XX DE Arabidopsis thaliana protein #56 modulated by PTGS.  
XX KW Posttranscriptional gene silencing; PTGS; plant; transformation.  
XX OS Arabidopsis thaliana.  
XX PN WO200281695-A2.  
XX PD 17-OCT-2002.  
XX PF 05-APR-2002; 2002WO-EP003806.  
XX PR 06-APR-2001; 2001US-0282049P.  
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PA (FRIE-) FRIEDRICH MIESCHER INST.  
XX PI Zhu T, Glazov EA, Meins F, Wang X, Chang H;  
XX DR WPI; 2003-103337/09.  
XX DR N-PSDB; ABZ42072.  
XX PT Novel polynucleic acid segment useful for modulating gene expression  
PT within a cell by posttranscriptional gene silencing, and for augmenting a  
PT plant cell genome.  
XX PS Claim 53; Page 273-274; 438pp; English.  
XX CC The invention relates to a novel isolated polynucleic acid segment  
CC modulated within a cell by posttranscriptional gene silencing (PTGS). The  
CC invention specifically relates to a method to identify an expression  
CC product that is modulated by PTGS. The polynucleotide is useful for  
CC modulating the gene expression within a cell by PTGS, by introducing the  
CC polynucleic acid into a cell and expressing the nucleic acid segment in  
CC the cell to form a product. The polynucleic acid segment is also useful  
CC for augmenting a cell genome, and for augmenting a plant genome, by  
CC contacting a plant cell with the segment to produce a transformed plant  
CC cell, and growing the transformed plant cell to produce a differentiated  
CC transformed plant. The sequences shown in ABP81173 - ABP81298 represent  
CC the product of a segment of A. thaliana cDNA modulated by PTGS  
XX SQ Sequence 324 AA;  
Query Match 41.1%; Score 46; DB 6; Length 324;  
Best Local Similarity 38.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
QY 1 RSDYKLYKNSSNSTLKNLG 21  
DB 14 RESFRLPKNGDGSITKKELG 34  
RESULT 26  
AAB18299  
ID AAB18299 standard; protein; 332 AA.  
XX AC AAB18299;  
XX DT 07-NOV-2000 (first entry)  
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:157.  
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX Plasmodium falciparum.  
PN WO2000025728-A2.  
XX 11-MAY-2000.  
XX PF 05-NOV-1999; 99WO-US026796.  
XX PR 05-NOV-1998; 98US-0107131P.  
XX (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
XX (VENT/) VENTER J C.  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection.  
XX PS Disclosure; Page 371-372; 577pp; English.  
XX CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II), (I) and  
CC (II) are useful for the development of vaccines against P. falciparum  
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
CC immunogens comprising the sequences of (I), are useful in the detection  
CC of infection with P. falciparum. Furthermore, (I) (especially when they  
CC are rifins or secreted or membrane proteins) can aid the identification  
CC of drugs to treat or prevent P. falciparum infection, or they can be used  
CC to identify drug resistance in P. falciparum. Sequencing of the  
CC Plasmodium chromosome 2 and the subsequent identification of proteins  
CC encoded by it will help to expand our understanding of parasite biology,  
CC a process hampered by the complexity of the parasitic lifecycle, and  
CC provide new targets for vaccine and drug development. Parasite resistance  
CC to drugs and mosquito resistance to insecticides have led to a resurgence  
CC of malaria in many parts of the world, and there is a pressing need for  
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
CC represent nucleotide and protein sequences given in the present  
CC invention, but which are not specifically mentioned within the  
CC specification  
XX SQ Sequence 332 AA;  
Query Match 41.1%; Score 46; DB 3; Length 332;  
Best Local Similarity 57.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 7 YKNKSSNSTLKNL 20  
DB 265 YNNNSNNNTSNNI 278  
RESULT 27  
ABU23678  
ID ABU23678 standard; protein; 2052 AA.  
XX AC ABU23678;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #9205.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Clostridium acetobutylicum.

```
XX WO200277183-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 21-MAR-2002; 2002WO-US009107.
XX
XX PR 21-MAR-2001; 2001US-00815242.
XX
XX PR 06-SEP-2001; 2001US-00948993.
XX
XX PR 25-OCT-2001; 2001US-0342923P.
XX
XX PR 08-FEB-2002; 2002US-00072851.
XX
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX DR WPI; 2003-029926/02.
XX
XX DR N-PSDB; ACA27548.
XX
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX PS Claim 25; SEQ ID NO 51602; 1766pp; English.
XX
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 613 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed, (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 2052 AA;
XX
XX Query Match 41.1%; Score 46; DB 6; Length 2052;
XX Best Local Similarity 55.6%; Pred. No. 1.1e+03;
XX Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 1;
XX
XX Qy 7 YNKNS--SNSLKNLGE 22
XX :|||||:|||||:
XX Db 1646 FEKNSNFSKNLKNIGD 1663
XX
XX RESULT 28
XX ADF04538
XX ID ADF04538 standard; protein; 60 AA.
```

```
XX ADF04538;
XX
XX AC 12-FEB-2004 (first entry)
XX
XX DT Bacterial polypeptide #651.
XX
XX DE Proteus mirabilis infection; bacterial infection; antibacterial;
XX KW immunostimulant.
XX
XX OS Proteus mirabilis.
XX
XX PN US6605709-B1.
XX
XX XX 12-AUG-2003.
XX
XX PD 05-APR-2000; 2000US-00543681.
XX
XX PF 09-APR-1999; 99US-0128706P.
XX
XX PR (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PA Breton GL;
XX
XX PI WPI; 2003-895291/82.
XX
XX DR N-PSDB; ADF00366.
XX
XX DR New Proteus mirabilis polypeptides and polynucleotides, useful as
XX PT reagents for diagnosis of bacterial disease, as components of
XX PT antibacterial vaccines, as targets for antibacterial drugs, or as
XX PT biocontrol agents for plants.
XX
XX PS Disclosure; SEQ ID NO 4823; 870pp; English.
XX
XX CC The invention relates to new Proteus mirabilis polypeptides and
XX CC polynucleotides. The invention also relates to antibodies against the
XX CC polypeptides, methods for producing the polypeptides, a method of
XX CC generating vaccines for immunising an individual against P. mirabilis, a
XX CC method for evaluating a compound for the ability to bind a P. mirabilis
XX CC polypeptide and a method for screening test compounds for anti-bacterial
XX CC activity. The polypeptides and polynucleotides are useful as molecular
XX CC targets for diagnosing, preventing and treating pathological conditions
XX CC resulting from bacterial infection, as reagents for diagnosis of
XX CC bacterial diseases, as components of antibacterial vaccines, as targets
XX CC for antibacterial drugs or as bio-control agents for plants. This
XX CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
XX SQ Sequence 60 AA;
XX
XX Query Match 40.2%; Score 45; DB 7; Length 60;
XX Best Local Similarity 56.2%; Pred. No. 23;
XX Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
XX
XX Qy 5 KLYNKNSSSNSTLKNL 20
XX :|||||:|||||:
XX Db 3 KLYKNSLTSTLTKRV 18
XX
XX RESULT 29
XX AAB62090
XX ID AAB62090 standard; protein; 225 AA.
XX
XX AC AAB62090;
XX
XX DT 29-MAY-2001 (first entry)
XX
XX DE T. indica Cox1 intronic open reading frame sequence.
XX
XX KW Intronic region: taxonomy; mitochondria; eukaryote; fungus; protozoa;
XX KW intron polymorphism analysis; cytochrome oxidase subunit 1; Cox1; Cox2;
XX KW Nadl.
XX
XX OS Tilletia indica.
```



```

XX PN WO200114592-A2.
XX PD 01-MAR-2001.
XX XX
XX PF 23-AUG-2000; 2000WO-US023119.
XX XX
XX PR 25-AUG-1999; 99US-0150977P.
XX XX
XX PA (CLAR-) CLARITY BIOSCIENCES INC.
XX PI Honeycutt R, McClelland M;
XX XX
XX DR WPI; 2001-218459/22.
XX DR N-PSDB; AAF57219.
XX XX
XX PT Constructing a profile of intronic regions, useful for characterizing
XX PT organisms e.g. eukaryotes and fungi, comprises selecting intronic regions
XX PT from nuclear or organellar gene sequences and analyzing the intronic
XX PT regions.
XX PS Example 8; Page 59; 69pp; English.
XX CC
XX CC The invention relates to constructing a profile of intronic regions (IRs)
XX CC from a given taxonomic group of organisms, useful for characterizing
XX CC target organisms. The method comprises (a) selecting an IR known to be
XX CC found in some or all members of the taxonomic group; (b) analyzing IR of
XX CC known members of the taxonomic group; and (c) constructing a profile of
XX CC IR characteristic from the taxonomic group. The target organism includes
XX CC eukaryotes, fungi, in particular Candida or Aspergillus and is suspected
XX CC of belonging to a smaller taxonomic group than the given taxonomic group.
XX CC Characterizing a target organism is useful for diagnosing the presence of
XX CC any organism that contains DNA including fungi, protozoan and other
XX CC members of the animal and plant kingdom. Fungi suitable for detection by
XX CC intron polymorphism analysis include Candida, Aspergillus, Cryptococcus,
XX CC Blastomyces for clinical applications and Fusarium, Tilletia and Botrytis
XX CC for non-clinical applications. The method is rapid and accurate in
XX CC detecting organisms. The present sequence represents an intronic open
XX CC reading frame sequence of a cytochrome oxidase subunit 1 (Cox1) intron
XX CC from Tilletia indica
XX SQ Sequence 225 AA;

Query Match 40.2%; Score 45; DB 4; Length 225;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 YKLYNKSSNSTLKNLGE 22
Db 182 YKTYTKKASSYTTWKEVSE 200

RESULT 30
AAE30847
ID AAE30847 standard; protein; 225 AA.
XX AC
XX AC AAE30847;
XX XX
XX XX 24-FEB-2003 (first entry)
XX XX
XX DE Tilletia indica cox1 IREP protein.
XX XX
XX XX Intronic region-encoded protein; IREP; antifungal agent; gene therapy;
XX XX cytochrome oxidase subunit 1; cox1.
XX XX
XX OS Tilletia indica.
XX XX
XX PN WO200277160-A2.
XX XX
XX PD 03-OCT-2002.
XX XX
XX PF 26-FEB-2002; 2002WO-US006069.
XX XX

01-MAR-2001; 2001US-0272801P.
(CLAR-) CLARITY BIOSCIENCES INC.
Honeycutt R, Ralph D;
WPI; 2003-029921/02.
N-PSDB; AAD48120.

Screening an agent for modulating cellular activity of a non-human
organism, useful for preparing antifungal agents.

Example 8; Page 57; 68pp; English.

The invention relates to a method of screening an agent which comprises a
nucleic acid coding for a protein that effects intronic region-encoded
protein (IREP)-mediated post-transcriptional processing of RNA. The
method involves providing the protein in an assay format adapted for
studying the effects of the protein on post-transcriptional processing of
pre-mRNA and assaying for the effects in the presence of the agent. The
method is useful for producing a composition comprising an agent that
modulates IREP-mediated post-transcriptional processing of RNA for
inhibiting growth of a non-human organism, particularly fungus, that is
associated with a host organism. The invention is also useful in gene
therapy. The present sequence is Tilletia indica cytochrome oxidase the
subunit 1 (cox1) IREP protein. This sequence is used to illustrate the
method of the invention

SQ Sequence 225 AA;

Query Match 40.2%; Score 45; DB 6; Length 225;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 YKLYNKSSNSTLKNLGE 22
Db 182 YKTYTKKASSYTTWKEVSE 200

RESULT 31
ABG93043
ID ABG93043 standard; protein; 338 AA.
XX AC
XX AC ABG93043;
XX XX
XX DT 21-NOV-2002 (first entry)
XX XX
XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 44.
XX XX
XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
XX KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
XX KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
XX KW neurodegeneration; cell death.
XX OS Saccharomyces cerevisiae.
XX XX
XX PN WO200264766-A2.
XX XX
XX PD 22-AUG-2002.
XX XX
XX PF 21-DEC-2001; 2001WO-EP015398.
XX XX
XX XX 22-DEC-2000; 2000EP-00870318.
XX PR 04-JAN-2001; 2001EP-00870002.
XX PR 09-JAN-2001; 2001EP-00870003.
XX XX
XX PA (JANC ) JANSSEN PHARM NV.
XX XX
XX XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX PI WPI; 2002-667002/71.
XX XX
XX DR N-PSDB; ABQ76309.
XX XX

```

PT New isolated nucleic acid representing a synthetic BAX-gene, useful as  
 PT medicament for treating, preventing and/or alleviating yeast or fungal  
 PT infections or proliferative disorders, or for preventing apoptosis in  
 PT certain diseases.  
 XX  
 XX  
 PS Claim 36; Fig 1; 344pp; English.  
 XX  
 CC This invention describes a novel nucleic acid representing a synthetic  
 CC Bax gene. The Bax gene of the invention is useful for identifying Bax-  
 CC resistant yeast or fungi, identifying, or obtaining and identifying  
 CC Candida spp. sequences that are differentially expressed in a pathway  
 CC eventually leading to programmed cell death or identifying inhibitors or  
 CC inhibitor sequences of Bax-induced cell death. The products of the  
 CC invention have cytostatic, fungicide; immunosuppressive, virucide and  
 CC vasotropic activity and can be used in vaccines or for gene therapy. The  
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
 CC antisense molecules and antibodies are useful as medicaments or in  
 CC preparing a medicament for treating, preventing and/or alleviating  
 CC diseases associated with yeast or fungi or proliferative disorders, such  
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
 CC or polypeptides, or the genetically modified organism are useful for  
 CC preparing a medicament for modifying the endogenous flora of humans and  
 CC other mammals. The vaccine is useful for immunising against yeast or  
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
 CC ischaemia, diseases related with viral infections or neurodegenerations.  
 CC This sequence represents a polypeptide associated with the Bax gene  
 CC described in the disclosure of the invention  
 XX  
 XX

SQ Sequence 338 AA;

Query Match 40.2%; Score 45; DB 5; Length 338;  
 Best Local Similarity 58.8%; Pred. No. 1.9e+02;

Matches 10; Conservative 3; Mismatches 0; Indels 4; Gaps 1;

Qy 7 YNKN-----SSNSTLKN 19

||||: |||: |||:

Db 109 YNKNKHVRSSNTTKN 125

RESULT 32

ABP28260

ID ABP28260 standard; protein; 423 AA.

AC

XX ABP28260;

XX

DT 02-JUL-2002 (first entry)

XX

DE Streptococcus polypeptide SEQ ID NO 5696.

XX

SW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW anti-inflammatory; infection; vaccine; meningitis; gene therapy.  
 XX

OS Streptococcus pyogenes.  
 XX

XX WO200234771-A2.  
 XX

XX 02-MAY-2002.  
 XX

XX 29-OCT-2001; 2001WO-GB004789.  
 XX

XX 27-OCT-2000; 2000GB-00026333.  
 XX

XX 24-NOV-2000; 2000GB-00028727.  
 XX

XX 07-MAR-2001; 2001GB-00005640.  
 XX

XX (CHIR-) CHIRON SPA.  
 XX

XX (GENO-) INST GENOMIC RES.  
 XX

XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
 XX Tettelin H;

XX WPI; 2002-352536/38.  
 XX

DR N-PSDB; ABN68891.

XX New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX

PS Claim 1; Page 3731; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX

SQ Sequence 423 AA;

Query Match 40.2%; Score 45; DB 5; Length 423;

Best Local Similarity 45.5%; Pred. No. 2.4e+02;

Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RSDYKLYNKNSSNSTLKNLGE 22

||||: |||: |||:

Db 92 RSDKRLKLNQVNSLKNMQE 113

RESULT 33

AAW20085

ID AAW20085 standard; protein; 509 AA.

XX

AC AAW20085;

XX

DT 29-MAY-1997 (first entry)

XX

DE Helicobacter pylori cytoplasmic protein, 10009666.aa.

XX

SW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis.  
 XX

OS Helicobacter pylori.  
 XX

XX WO9640893-A1.  
 XX

XX 19-DEC-1996.  
 XX

XX 06-JUN-1996; 96WO-US009122.  
 XX

XX 07-JUN-1995; 95US-00487032.  
 XX

XX 01-APR-1996; 96US-00630405.  
 XX

XX (ASTR ) ASTRA AB.  
 XX

XX Smith D, Berglindh OT, Mellgaerd BL;  
 XX

XX WPI; 1997-052306/05.  
 XX

XX N-PSDB; AAT67337.  
 XX

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX

PS Claim 61; Page 321-322; 1481pp; English.

XX The present sequence is a Helicobacter pylori protein, which may be used  
 CC in a vaccine to prevent or treat H. pylori infection or to identify H.  
 CC pylori polypeptide binding compounds, useful as potential H. pylori life  
 CC cycle activators or inhibitors. The genomic sequence of H. pylori (ATCC  
 CC 55679) was determined from overlapping contigs generated by mechanically  
 CC shearing the bacterial DNA. The sequences were analysed for ORF of at  
 CC least 180 nucleotides, and the predicted coding regions defined by  
 CC computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts  
 XX

SO Sequence 509 AA;  
 Query Match 40.2%; Score 45; DB 2; Length 509;  
 Best Local Similarity 58.8%; Pred. No. 3e+02;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 YKLYNKNSSNSTLKNL 20  
 | | | | | | | | | |  
 Db 296 YALVGKNASGKSTLNL 312

RESULT 34  
 ID AAW20918 standard; protein; 593 AA.  
 XX AAW20918;  
 XX 21-JUL-1997 (first entry)  
 DT  
 XX H. pylori transporter protein, 14gp12015orf14.  
 XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX Helicobacter pylori.  
 XX WO9640893-A1.  
 XX 19-DEC-1996.  
 XX 06-JUN-1996; 96WO-US009122.  
 XX 07-JUN-1995; 95US-00487032.  
 PR 01-APR-1996; 96US-00630405.  
 XX (ASTR ) ASTRA AB.  
 XX Smith D, Berglindh OT, Mellgaard BU;  
 XX WPI; 1997-052306/05.  
 DR N-PSDB; AAT68171.  
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX Claim 56; Page 1317-1318; 1481pp; English.  
 XX The present sequence is a H. pylori transporter protein. The protein may  
 CC be used in a vaccine to prevent or treat H. pylori infection or to  
 CC identify H. pylori polypeptide binding compounds, useful as potential H.  
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.  
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed for  
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined  
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were

CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts  
 XX

SO Sequence 593 AA;  
 Query Match 40.2%; Score 45; DB 2; Length 593;  
 Best Local Similarity 58.8%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 YKLYNKNSSNSTLKNL 20  
 | | | | | | | | | |  
 Db 380 YALVGKNASGKSTLNL 396

RESULT 35  
 ID ABG14461 standard; protein; 657 AA.  
 XX ABG14461;  
 XX 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #14452.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS78648.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 44820; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 657 AA;

Query Match 40.2%; Score 45; DB 4; Length 657;  
Best Local Similarity 60.0%; Pred. No. 4.1e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels

Qy 1 RSDYKLYNKSSNS 15  
|:|:|:|:|:|  
Db 536 RTDYQLYASGSSRS 550

RESULT 36  
ADA34312  
ID ADA34312 standard; protein; 733 AA.

ADA34312;

DT 20-NOV-2003 (first entry)

DE *Acinetobacter baumannii* protein #1473.

JW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
 KW plant biocontrol agent.  
 KW plant biocontrol agent.

OS Acinetobacter baumannii.

PN US6562958-B1.

PD 13-MAY-2003.

04-JUN-1999; 99US-00328352.

PR 09-JUN-1998: 98US-0088701P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton G, Bush D;

DR WPI; 2003-576092/54.

XX  
XX  
: 0070 CURRY, JENNIFER A

New *Acinetobacter baumannii* proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Example: SEO ID NO 5599: 328pp: English.

CC The invention relates to isolated *Acinetobacter baumannii* nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other *Acinetobacter* species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents the amino acid sequence of an A.  
CC baumannii protein.

Sequence 733 AA:

Query Match 40.2%; Score 45; DB 6; Length 733;  
Best Local Similarity 41.2%; Pred. No. 4.7e+02;  
Matches 7; Conservative 6; Mismatches 4; Indels

QY 1 RSDYKLYNKNSSSNSTL 17

[illegible]

RESULT 37

|           |   |                         |  |
|-----------|---|-------------------------|--|
| PH        | Key   | Location/Qualifiers     |  |
| FT        | Misc-difference 586   | /note= "encoded by RTC" |  |
| XX        |   |                         |  |
| XX        | WO200136469-A2.   |                         |  |
| XX        | 25-MAY-2001.  |                         |  |
| XX        |   |                         |  |
| XX        | 16-NOV-2000; 2000WO-US031472.   |                         |  |
| XX        |   |                         |  |
| XX        | 19-NOV-1999; 99US-0166429P.   |                         |  |
| XX        | (LEXI-) LEXICON GENETICS INC.   |                         |  |
| PA        | Turner CA, Mathur B;  |                         |  |
| PI        |   |                         |  |
| XX        |   |                         |  |
| DR        | WPI; 2001-355614/37.  |                         |  |
| DR        | N-PSDB; AAF83849.   |                         |  |
| XX        |   |                         |  |
| PT        | New polynucleotides encoding human secreted proteins that share           |                         |  |
| PT        | structural similarity with animal ceruloplasmins, useful for drug         |                         |  |
| PT        | screening, diagnosis and treatment of biological disorders or imbalances. |                         |  |
| XX        |   |                         |  |
| PS        | Claim 2; Page 26-29; 28pp; English.                                       |                         |  |
| XX        |   |                         |  |
| CC        | This represents a novel human secreted protein (NHP) that share           |                         |  |
| CC        | structural similarity with animal ceruloplasmins. NHP oligonucleotides    |                         |  |
| CC        | are useful as hybridization probes for screening libraries and assessing  |                         |  |
| CC        | gene expression patterns. Sequences derived from regions adjacent to the  |                         |  |
| CC        | intron/exon boundaries of NHP gene can be used to design primers for use  |                         |  |
| CC        | in amplification assays to detect mutations within the exons, splice      |                         |  |
| CC        | sites, introns that can be used in diagnostics and pharmacogenomics. NHP  |                         |  |
| CC        | nucleotide sequences are useful for drug screening and nucleotide         |                         |  |
| CC        | construct encoding NHP products are useful in gene therapy for modulating |                         |  |
| CC        | NHP expression and to produce genetically engineered host cells to        |                         |  |
| CC        | express NHP products in vivo. NHP products e.g. NHP fusion protein        |                         |  |
| CC        | products, NHP antibodies and antagonists or agonists are useful as        |                         |  |
| CC        | therapeutics i.e., for the treatment of Wilson's disease. The encoded NHP |                         |  |
| CC        | polypeptides are useful for generating antibodies, as reagents in         |                         |  |
| CC        | diagnostic assays, for identifying other cellular gene products related   |                         |  |
| CC        | to NHP and as reagents in assays for screening for compounds that are     |                         |  |
| CC        | useful in the treatment of mental, biological or medical disorders and    |                         |  |
| CC        | diseases  |                         |  |
| XX        |   |                         |  |
| SQ        | Sequence 996 AA;  |                         |  |
|           | Query Match 40.2%; Score 45; DB 4; Length 996;                            |                         |  |
|           | Best Local Similarity 42.1%; Pred. No. 6.8e+02;                           |                         |  |
|           | Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;                |                         |  |
| Qy        | 4 YKLYNKNSSNSTLKNLGE 22   |                         |  |
| Db        | 774 YRQYDDNTFTQTKRNEGE 792  |                         |  |
|           | :   :   :   :   :   |                         |  |
| RESULT 39 |   |                         |  |
| ABG69624  |   |                         |  |
| ID        | ABG69624 standard; protein; 1049 AA.                                      |                         |  |
| XX        |   |                         |  |
| AC        | ABG69624;   |                         |  |
| XX        |   |                         |  |
| DT        | 21-OCT-2002 (first entry)   |                         |  |
| XX        |   |                         |  |
| DE        | Human secreted protein SCEP-4.  |                         |  |
| XX        |   |                         |  |
| KW        | Secreted protein; SCEP; human; cell proliferative disorder; cancer;       |                         |  |
| KW        | keratosis; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;       |                         |  |
| KW        | psoriasis; autoimmune disorder; inflammatory disorder; AIDS; arthritis;   |                         |  |
| KW        | acquired immunodeficiency syndrome; adult respiratory distress syndrome;  |                         |  |
| KW        | Addison's disease; allergy; asthma; osteoporosis; autoimmune thyroiditis; |                         |  |
| KW        | Crohn's disease; dermatitis; diabetes; Graves' disease; haemodialysis;    |                         |  |
| KW        | glomerulonephritis; scleroderma; systemic lupus erythematosus; uveitis;   |                         |  |
| KW        | systemic sclerosis; ulcerative colitis; infection; trauma; Pick disease;  |                         |  |
| KW        | cardiovascular disorder; congestive heart failure; angina; epilepsy;      |                         |  |
| KW        | hypertensive heart disease; neurological disorder; Alzheimer's disease;   |                         |  |
| KW        | Parkinson's disease; amyotrophic lateral sclerosis; stroke; dementia;     |                         |  |
| KW        | Huntington's disease; multiple sclerosis; neuromuscular disorder;         |                         |  |
| KW        | metabolic disorder; endocrine disorder; toxic myopathy; mental disorder;  |                         |  |
| KW        | schizophrenic disorder; developmental disorder; anaemia; epilepsy;        |                         |  |
| KW        | hypothyroidism; glaucoma; sensorineural hearing loss; cataract;           |                         |  |
| XX        | transgenic animal.  |                         |  |
| XX        |   |                         |  |
| OS        | Homo sapiens.   |                         |  |
| XX        |   |                         |  |
| PN        | WO200248337-A2.   |                         |  |
| XX        |   |                         |  |
| PD        | 20-JUN-2002.  |                         |  |
| XX        |   |                         |  |
| PP        | 12-DEC-2001; 2001WO-US048517.   |                         |  |
| XX        |   |                         |  |
| PR        | 13-DEC-2000; 2000US-0255639P.   |                         |  |
| PR        | 21-DEC-2000; 2000US-0257852P.   |                         |  |
| PR        | 05-JAN-2001; 2001US-0260105P.   |                         |  |
| PR        | 18-JAN-2001; 2001US-0262932P.   |                         |  |
| PR        | 18-JAN-2001; 2001US-0263096P.   |                         |  |
| PR        | 19-JAN-2001; 2001US-0263090P.   |                         |  |
| PR        | 02-FEB-2001; 2001US-0265926P.   |                         |  |
| XX        |   |                         |  |
| PA        | (INCY-) INCYTE GENOMICS INC.  |                         |  |
| XX        |   |                         |  |
| PI        | Griffin JA, Yao MG, Duggan BM, Yue H, Ding L, Lal PG, Lee EA;             |                         |  |
| PI        | Ramkumar J, Thangavelu K, Xu Y, Lee S, Tang YT, Nguyen DB;                |                         |  |
| PI        | Warren BA, Honchell CD, Gietzen KJ, Baughn MR, Gandhi AR, Arvizu C;       |                         |  |
| PI        | Walia NK, Lu Y, Elliott VM, Lu DAM, Hafalia AJA, Azimzai Y,               |                         |  |
| PI        | Khan FA, Tran UK;   |                         |  |
| DR        |   |                         |  |
| XX        | WPI; 2002-583509/62.  |                         |  |
| DR        | N-PSDB; ABK99921.   |                         |  |
| XX        |   |                         |  |
| PT        | Novel human secreted proteins and polynucleotides for diagnosing,         |                         |  |
| PT        | preventing or treating cell proliferative, autoimmune/inflammatory,       |                         |  |
| PT        | cardiovascular, neurological and developmental disorders.                 |                         |  |
| XX        |   |                         |  |
| PS        | Claim 1; Page 164-166; 234pp; English.                                    |                         |  |
| XX        |   |                         |  |
| CC        | The invention describes an isolated polypeptide chosen from secreted      |                         |  |
| CC        | proteins (I), SCEP 1-54. (I) and the polynucleotide encoding it (II) are  |                         |  |
| CC        | useful for screening a compound for effectiveness as an agonist or        |                         |  |
| CC        | antagonist of (I) or compound that alters expression of (II). (I), the    |                         |  |
| CC        | identified agonist and antagonist are useful for treating a disease or    |                         |  |
| CC        | condition associated altered expression of functional SCEP in a patient.  |                         |  |
| CC        | An antibody specific to (I) is useful for detecting the presence of (I),  |                         |  |
| CC        | purifying (I) from a sample and for diagnosing a condition or disease     |                         |  |
| CC        | associated with expression of SCEP in a subject or in a biological        |                         |  |
| CC        | sample. (I) and (II) and modulators of (I) are useful for diagnosis,      |                         |  |
| CC        | treatment and prevention of cell proliferative disorders (e.g. cancer,    |                         |  |
| CC        | keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and    |                         |  |
| CC        | psoriasis), autoimmune/inflammatory disorders (e.g. acquired              |                         |  |
| CC        | immunodeficiency syndrome (AIDS), adult respiratory distress syndrome,    |                         |  |
| CC        | Addison's disease, allergies, asthma, osteoporosis, autoimmune            |                         |  |
| CC        | thyroiditis, Crohn's disease, dermatitis, diabetes, Graves' disease,      |                         |  |
| CC        | glomerulonephritis, arthritis, scleroderma, systemic lupus erythematosus, |                         |  |
| CC        | systemic sclerosis, ulcerative colitis, haemodialysis, uveitis; viral,    |                         |  |
| CC        | bacterial, fungal, parasitic, protozoal, helminthic infections and        |                         |  |
| CC        | trauma), cardiovascular disorders (e.g. congestive heart failure, angina, |                         |  |
| CC        | hypertensive heart disease), neurological disorders (e.g. Alzheimer's and |                         |  |
| CC        | Pick disease, Parkinson disease, amyotrophic lateral sclerosis, epilepsy, |                         |  |
| CC        | stroke, Huntington's disease, multiple sclerosis, dementia, neuromuscular |                         |  |
| CC        | disorders, metabolic, endocrine and toxic myopathies, mental disorders,   |                         |  |
| CC        | schizophrenic disorders, and developmental disorders (e.g. anaemia,       |                         |  |
| CC        | epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and        |                         |  |
| CC        | cataract). (II) is useful for creating transgenic animals to model human  |                         |  |
| CC        | disease and to detect and quantify gene expression in biopsied tissues in |                         |  |
| CC        | which expression of SCEP is correlated with disease. This is the amino    |                         |  |
| CC        | acid sequence of a human secreted protein (SCEP)                          |                         |  |
| XX        |   |                         |  |

Job time : 98.8913 secs

```

SQ      Sequence 1049 AA;

Query Match      40.2%; Score 45; DB 5; Length 1049;
Best Local Similarity 42.1%; Pred.No. 7.2e+02;
Matches      8; Conservative      5; Mismatches      6; Indels      0; Gaps      0;

QY      4 YKLYNKSSNSNLTKNLGE 22
      |:|:|:|:|:|
Db      771 YRQYDDNTFTNQTNRNEGE 789

RESULT 40
ABR52794
ID      ABR52794 standard; protein; 1758 AA.
XX
AC      ABR52794;
XX
DT      20-JUN-2003 (first entry)
XX
DE      Protein sequence #SEQ ID 453.
XX
KW      Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS      Saccharomyces cerevisiae.
XX
PN      EP1258494-A1.
XX
PD      20-NOV-2002.
XX
PF      20-DEC-2001; 2001EP-00130253.
XX
PR      15-MAY-2001; 2001EP-00111774.
XX
PC      (CELL-) CELLZONE AG.
XX
PI      Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI      Marzioch M, Schultz JD, Superti-Furga GD;
DR      WPI; 2003-250078/25.
DR      N-PSDB; ACC50836.
XX
XX      New isolated protein complexes useful for diagnosing a disease or
PT      disorder, or as a target for an active agent of a pharmaceutical,
PT      preferably a drug target in the treatment or prevention of disease or
PT      disorder.
XX
PS      Disclosure; SEQ ID NO 453; 17pp + Sequence Listing; English.
XX
XX      The invention relates to multiprotein complexes from eukaryotes. Proteins
CC      of the invention and DNA sequences encoding them are given in records
CC      ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC      obtainable by using a protein as a bait and isolating the set of proteins
CC      which is attached thereto from cells. Such protein complexes may comprise
CC      up to 30 distinct proteins. Protein complexes of the invention are useful
CC      for diagnosing a disease or disorder, or as a target for an active agent
CC      of a pharmaceutical, preferably a drug target in the treatment or
CC      prevention of a disease or disorder. Note: The sequence data for this
CC      patent is not represented in the printed specification, but is based on
CC      sequence information supplied by the European Patent Office. The complete
CC      document is available on CD-ROM
XX
SQ      Sequence 1758 AA;

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 24, 2004, 09:07:49 ; Search time 56.663 Seconds  
(without alignments)  
82.302 Million cell updates/sec

Title: US-09-719-379A-3

Perfect score: 72

Sequence: 1 RSDYKFDYDKRID 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID      | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| 1          | 72    | 100.0       | 13     | 3 AAY79981 | Aay79981 Non-typea |
| 2          | 72    | 100.0       | 13     | 3 AAY79989 | Aay79989 Non-typea |
| 3          | 72    | 100.0       | 13     | 3 AAY79985 | Aay79985 Non-typea |
| 4          | 72    | 100.0       | 13     | 4 AAB47441 | Aab47441 LB1(f) co |
| 5          | 72    | 100.0       | 22     | 4 AAB47445 | Aab47445 Entire 3r |
| 6          | 72    | 100.0       | 464    | 3 AAY79993 | Aay79993 Plasmid L |
| 7          | 47    | 65.3        | 19     | 3 AAY79957 | Aay79957 Non-typea |
| 8          | 44    | 61.1        | 19     | 3 AAY79958 | Aay79958 Non-typea |
| 9          | 43.5  | 60.4        | 19     | 3 AAY79968 | Aay79968 Non-typea |
| 10         | 43    | 59.7        | 18     | 2 AAW67572 | Aaw67572 Non-typea |
| 11         | 43    | 59.7        | 18     | 6 ADA25163 | Ada25163 H. influe |
| 12         | 43    | 59.7        | 18     | 7 ADC89652 | Adc89652 H. influe |
| 13         | 43    | 59.7        | 19     | 3 AAY79982 | Aay79982 Non-typea |
| 14         | 43    | 59.7        | 19     | 3 AAY79969 | Aay79969 Non-typea |
| 15         | 43    | 59.7        | 19     | 3 AAY79991 | Aay79991 Non-typea |
| 16         | 43    | 59.7        | 19     | 3 AAY79965 | Aay79965 Non-typea |
| 17         | 43    | 59.7        | 19     | 3 AAY79956 | Aay79956 Non-typea |
| 18         | 43    | 59.7        | 19     | 3 AAY79964 | Aay79964 Non-typea |
| 19         | 43    | 59.7        | 19     | 3 AAY79955 | Aay79955 Non-typea |
| 20         | 43    | 59.7        | 19     | 4 AAB47439 | Aab47439 LB1(f) co |
| 21         | 43    | 59.7        | 20     | 3 AAB20881 | Aab20881 LB1gr1 pe |
| 22         | 43    | 59.7        | 28     | 4 AAB47443 | Aab47443 Entire 3r |
| 23         | 43    | 59.7        | 40     | 2 AAW67581 | Aaw67581 Synthetic |
| 24         | 43    | 59.7        | 40     | 3 AAY79986 | Aay79986 Measles v |
| 25         | 43    | 59.7        | 40     | 6 ADA25172 | Ada25172 Chimeric  |

#### ALIGNMENTS

##### RESULT 1

AAY79981

ID AAY79981 standard; peptide; 13 AA.

AC AAY79981;

DT 15-MAY-2000 (first entry)

DE Non-typeable H. influenzae group 3 LB1(f) peptide N1729MEE.

KW Vaccine; non-typeable Haemophilus influenzae; ntlH; infection;  
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
KW conjunctivitis; lower respiratory tract infection.

OS Haemophilus influenzae.

PN WO9964067-A2.

PD 16-DEC-1999.

PF 28-MAY-1999; 99WO-US011980.

PR 11-JUN-1998; 98GB-00012613.

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
(OHIS ) UNIV OHIO STATE RES FOUND.

PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

DR Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
XX Haemophilus influenza.

FS Example 1; Page 30; 69pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
CC prevention, and treatment of Haemophilus influenzae infections such as  
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
CC infection. The peptides may also be used in vaccines against H.  
CC influenza. Antibodies and probes from the present invention can be used  
CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
CC the present invention

XX

```

SQ Sequence 13 AA;
  Query Match      100.0%; Score 72; DB 3; Length 13;
  Best Local Similarity 100.0%; Pred. No. 6.6e-06;
  Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13
Db 1 RSDYKFYDNKRID 13

RESULT 2
AAAY79989
ID AAAY79989 standard; peptide; 13 AA.
XX
AC AAAY79989;
XX
DT 15-MAY-2000 (first entry)
XX
DE Non-typeable H. influenzae 1729MEE Group 3 type peptide.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX
FN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
FN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
PT Haemophilus influenza.
XX
PS Claim 4; Page 46; 68pp; English.
XX
CC The present invention describes antigenic P5-like fimbria subunit
CC peptides (Lb1(f) peptides) of P5-like fimbria proteins from various
CC Haemophilus influenzae strains. The peptides are used for diagnosis,
CC prevention, and treatment of Haemophilus influenzae infections, such as
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
CC infection. The peptides may also be used in vaccines against H.
CC influenzae. Antibodies and probes from the present invention can be used
CC for diagnosis of H. influenzae infection. AAAY79955 to AAAY79993, and
CC AAAY791201 to AAAY791252, represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 13 AA;
  Query Match      100.0%; Score 72; DB 3; Length 13;
  Best Local Similarity 100.0%; Pred. No. 6.6e-06;
  Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDNKRID 13
Db 1 RSDYKFYDNKRID 13

RESULT 4
AAB47441
ID AAB47441 standard; peptide; 13 AA.
XX
AC AAB47441;
XX
DT 31-OCT-2001 (first entry)
XX
DE Lb1(f) cont. peptide from strain nH1-1729MEE (Group 3 type).
XX
KW surface exposed loop; major outer membrane protein P5; MOMP P5;
KW non-typeable H. influenzae; nH1; Lb1(f) peptide; B cell epitope;
KW otitis media; sinusitis; conjunctivitis;
KW lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX

```



```

PN WO200161013-A1.
PD 23-AUG-2001.
XX
XX 13-FEB-2001; 2001WO-EP001556.
XX
XX 15-FEB-2000; 2000GB-00003502.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Denoel P, Poolman J, Thonnard J;
XX WPI; 2001-522599/57.
XX
XX Recombinant bacterial outer membrane protein where one or more surface-
XX exposed loops are modified is useful as a vaccine to prevent or treat
XX Haemophilus influenzae infection or associated disease, e.g., otitis
XX media and conjunctivitis.
XX
XX Claim 1; Page 26; 29pp; English.
XX
XX The sequences given in AAB47439-46 represent peptides which may be used
XX to replace one or more surface exposed loops of major outer membrane
XX protein P5 (MOMP P5) of non-typeable H. influenzae (nHi). Each of these
XX peptides contain an LBI(f) peptide which is a 19 amino acid peptide
XX derived from the sequence of MOMP P5 from strain nHil128, representing
XX amino acids Arg117 to Gly135. This peptide represents the third exposed
XX loop of P5 and is a potential B cell epitope. The loops of the invention
XX are modified in terms of being in a non-native environment in the
XX recombinant outer membrane protein. The modified MOMP P5 may be used to
XX induce an immune response in a mammal to prevent or treat Haemophilus
XX influenzae infection or associated disease, e.g., otitis media,
XX sinusitis, conjunctivitis, or lower respiratory tract infection
XX
XX Sequence 13 AA;
XX
XX Query Match 100.0%; Score 72; DB 4; Length 13;
XX Best Local Similarity 100.0%; Pred. No. 6.6e-06;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RSDYKFYDNKRID 13
XX | | | | | | | | | |
XX DB 1 RSDYKFYDNKRID 13
XX
XX RESULT 5
XX AAB47445
XX ID AAB47445 standard; peptide; 22 AA.
XX
XX AC AAB47445;
XX
XX 31-OCT-2001 (first entry)
XX
XX Entire 3rd loop from strain nHi-1729WEE (Group 3 type).
XX
XX surface exposed loop; major outer membrane protein P5; MOMP P5;
XX non-typeable H. influenzae; nHi; LBI(f) peptide; B cell epitope;
XX otitis media; sinusitis; conjunctivitis;
XX lower respiratory tract infection.
XX
XX Haemophilus influenzae.
XX
XX WO200161013-A1.
XX
XX 23-AUG-2001.
XX
XX 13-FEB-2001; 2001WO-EP001556.
XX
XX 15-FEB-2000; 2000GB-00003502.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Denoel P, Poolman J, Thonnard J;
XX WPI; 2001-522599/57.
XX
XX Recombinant bacterial outer membrane protein where one or more surface-
XX exposed loops are modified is useful as a vaccine to prevent or treat
XX Haemophilus influenzae infection or associated disease, e.g., otitis
XX media and conjunctivitis.
XX
XX Claim 2; Page 26; 29pp; English.
XX
XX The sequences given in AAB47439-46 represent peptides which may be used
XX to replace one or more surface exposed loops of major outer membrane
XX protein P5 (MOMP P5) of non-typeable H. influenzae (nHi). Each of these
XX peptides contain an LBI(f) peptide which is a 19 amino acid peptide
XX derived from the sequence of MOMP P5 from strain nHil128, representing
XX amino acids Arg117 to Gly135. This peptide represents the third exposed
XX loop of P5 and is a potential B cell epitope. The loops of the invention
XX are modified in terms of being in a non-native environment in the
XX recombinant outer membrane protein. The modified MOMP P5 may be used to
XX induce an immune response in a mammal to prevent or treat Haemophilus
XX influenzae infection or associated disease, e.g., otitis media,
XX sinusitis, conjunctivitis, or lower respiratory tract infection
XX
XX Sequence 22 AA;
XX
XX Query Match 100.0%; Score 72; DB 4; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-05;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RSDYKFYDNKRID 13
XX | | | | | | | | | |
XX DB 3 RSDYKFYDNKRID 15
XX
XX RESULT 6
XX AAY79993
XX ID AAY79993 standard; protein; 464 AA.
XX
XX AC AAY79993;
XX
XX 15-MAY-2000 (first entry)
XX
XX Plasmid LPD-LBI-III protein sequence.
XX
XX Vaccine; non-typeable Haemophilus influenzae; nHi; infection;
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;
XX conjunctivitis; lower respiratory tract infection.
XX
XX Haemophilus influenzae.
XX
XX Synthetic.
XX
XX WO9964067-A2.
XX
XX 16-DEC-1999.
XX
XX 28-MAY-1999; 99WO-US011980.
XX
XX 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX N-FSDB; AAZ91252.
XX
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenzae.
XX
XX Claim 14; Fig 5; 68pp; English.
XX
XX

```

CC The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 464 AA;

Query Match 100.0%; Score 72; DB 3; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 0.00036;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFPYDNKRID 13  
 |||||  
 Db 442 RSDYKFPYDNKRID 454

## RESULT 7

AA79957  
 ID AAY79957 standard; peptide; 19 AA.  
 AC AAY79957;

DT 15-MAY-2000 (first entry)

DE Non-typeable H. influenzae group 1 LB1(f) peptide N1234MEE.

XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX

OS Haemophilus influenzae.

XX  
 PN WO9964067-A2.  
 XX

PD 16-DEC-1999.

PF 28-MAY-1999; 99WO-US011980.

PR 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

PI WPI; 2000-116457/10.

DR Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX

PS Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX

SQ Sequence 19 AA;

Query Match 65.3%; Score 47; DB 3; Length 19;

Best Local Similarity 88.9%; Pred. No. 0.28;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFPYDN 9  
 |||||  
 Db 1 RSDYKFPYDD 9

## RESULT 8

AA79958  
 ID AAY79958 standard; peptide; 19 AA.  
 XX

AC AAY79958;

XX 15-MAY-2000 (first entry)

XX Non-typeable H. influenzae group 1 LB1(f) peptide N90100RM.

XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX

OS Haemophilus influenzae.

XX  
 PN WO9964067-A2.  
 XX

PD 16-DEC-1999.

PF 28-MAY-1999; 99WO-US011980.

PR 11-JUN-1998; 98GB-00012613.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 PI WPI; 2000-116457/10.  
 XX

DR Novel antigenic P5-like fimbrin subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX

PS Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbrin subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX

SQ Sequence 19 AA;

Query Match 61.1%; Score 44; DB 3; Length 19;  
 Best Local Similarity 70.0%; Pred. No. 0.97;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFPYDNK 10  
 |||||  
 Db 1 RSDYKFPYDE 10

## RESULT 9

AA79968  
 ID AAY79968 standard; peptide; 19 AA.  
 XX

AC AAY79968;

XX



```

Query Match      59.7%; Score 43; DB 6; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.4;
Matches 7; Conservative 2; Mismatches 0; Gaps 0;

Qy 1 RSDYKFYDN 9
Db 1 RSDYKFYED 9

RESULT 12
ADC89652
ID ADC89652 standard; peptide; 18 AA.
XX AC
XX ADC89652;
XX DT 01-JAN-2004 (first entry)
XX DE H. influenzae fimbrin peptide #1.
XX KW Fimbrin; T cell epitope; vaccine; otitis media; auditory;
XX KW antiinflammatory.
XX OS Haemophilus influenzae.
XX PN US2003113344-A1.
XX PD 19-JUN-2003.
XX PF 19-AUG-2002; 2002US-00223711.
XX PR 04-SEP-1998; 98US-00148711.
XX PA (BAKA/) BAKALETZ L O.
XX PA (KAUM/) KAUMAYA P T P.
XX PI Bakaletz LO, Kaumaya PTP;
XX DR WPI; 2003-810881/76.
XX PT Novel synthetic chimeric fimbrin peptide LB1 or LB2 comprising a first
XX PT peptide unit, T cell epitope as second peptide unit and third linker
XX PT peptide unit, useful for preventing or reducing severity of otitis media.
XX PS Claim 1; SEQ ID NO 1; 15pp; English.
XX CC The invention relates to a synthetic chimaeric fimbrin peptide LB1 or LB2
XX CC comprises a first peptide unit derived from H. influenzae fimbrin , a
XX CC second peptide unit containing a T cell epitope and a third linker
XX CC peptide which connects the first peptide to the second. The chimaeric
XX CC peptide is useful for inducing an immune response in animals against non-
XX CC typable Haemophilus influenzae (NTHi) and for preventing or reducing
XX CC adherence of NTHi to host cells thereby preventing or reducing the
XX CC severity of otitis media. The present sequence is an H. influenzae
XX CC fimbrin peptide for use in the chimaeric peptides of the invention.
XX SQ Sequence 18 AA;

Query Match      59.7%; Score 43; DB 7; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9
Db 1 RSDYKFYED 9

RESULT 13
AAV79982
ID AAV79982 standard; peptide; 19 AA.
XX AC
XX AAV79982;

Query Match      59.7%; Score 43; DB 3; Length 19;
Best Local Similarity 77.8%; Pred. No. 1.5;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9
Db 1 RSDYKFYED 9

RESULT 14
AAV79969
ID AAV79969 standard; peptide; 19 AA.
XX AC
XX AAV79969;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 1 LB1(f) peptide NTHi-480.
XX KW Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.

```

```

DT 15-MAY-2000 (first entry)
XX Non-typeable H. influenzae group 1 LB1(f) peptide N1128.
XX KW Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.
XX PF 28-MAY-1999; 99WO-US011980.
XX PR 11-JUN-1998; 98GB-00012613.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX DR WPI; 2000-116457/10.
XX PT Novel antigenic P5-like fimbrin subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PS Example 1; Page 30; 68pp; English.
XX CC The present invention describes antigenic P5-like fimbrin subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAV79955 to AAV79993, and
XX CC AA291201 to AA291252, represent sequences used in the exemplification of
XX CC the present invention
XX SQ Sequence 19 AA;

Query Match      59.7%; Score 43; DB 3; Length 19;
Best Local Similarity 77.8%; Pred. No. 1.5;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9
Db 1 RSDYKFYED 9

RESULT 14
AAV79969
ID AAV79969 standard; peptide; 19 AA.
XX AC
XX AAV79969;
XX DT 15-MAY-2000 (first entry)
XX DE Non-typeable H. influenzae group 1 LB1(f) peptide NTHi-480.
XX KW Vaccine; non-typeable Haemophilus influenzae; nTHi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS Haemophilus influenzae.
XX PN WO9964067-A2.
XX PD 16-DEC-1999.

```

|  |
|--|
| XX   |
| Pf 28-MAY-1999;    99WO-US011980.  |
| XX   |
| PR 11-JUN-1998;    98GB-00012613.  |
| XX   |
| PA (SMIK ) SMITKLINE BEECHAM BIOLOGICALS.  |
| PA (OHIS ) UNIV OHIO STATE RES FOUND.  |
| XX   |
| PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  |
| XX WPI; 2000-116457/10.  |
| DR   |
| XX Novel antigenic P5-like fimbriae subunit peptides used in vaccines against  |
| PT Haemophilus influenza.  |
| XX Example 1; Page 29; 68pp; English.  |
| CC The present invention describes antigenic P5-like fimbriae subunit  |
| CC peptides (LBI(f) peptides) of P5-like fimbriae proteins from various  |
| CC Haemophilus influenza strains. The peptides are used for diagnosis,   |
| CC prevention, and treatment of Haemophilus influenza infections, such as  |
| CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract   |
| CC infection. The peptides may also be used in vaccines against H.   |
| CC influenza. Antibodies and probes from the present invention can be used   |
| CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and   |
| CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  |
| CC the present invention   |
| XX Sequence 19 AA;   |
| SQ   |
| Query Match                 59.7%; Score 43; DB 3; Length 19;  |
| Best Local Similarity      77.8%; Pred. No. 1.5;   |
| Matches          7; Conservative    2; Mismatches    0; Indels    0; Gaps    0   |
| Qy        1 RSDYKFVDN 9<br> <br>Db         <br>:<br>:<br>:<br>:<br>:<br>:<br>:<br>:<br>:   |
| Dd        1 RSDYKFVED 9  |
| RESULT 15  |
| AAY79991   |
| ID AAY79991 standard; peptide; 19 AA.  |
| AC AC  |
| AA79991;   |
| DT 15-MAY-2000 (first entry)   |
| DE Non-typeable H. influenzae l128 Group 1 type peptide.   |
| KW Vaccine; non-typeable Haemophilus influenzae; ntHi; infection;<br>KW chimeric protein; Haemophilus influenzae; P5-like fimbriae protein;<br>KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;<br>KW conjunctivitis; lower respiratory tract infection.<br>OS Haemophilus influenzae.<br>PN WO9964067-A2.<br>PD Best DEC-1999.<br>PF 28-MAY-1999;    99WO-US011980.<br>PP 11-JUN-1998;    98GB-00012613.<br>PR (SMIK ) SMITKLINE BEECHAM BIOLOGICALS.<br>PA (OHIS ) UNIV OHIO STATE RES FOUND.<br>XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y; |
| PI WPI; 2000-116457/10.  |
| DR   |
| XX Novel antigenic P5-like fimbriae subunit peptides used in vaccines against  |
| PT Haemophilus influenza.  |

Claim 11; Page 46; 68pp; English.

The present invention describes antigenic P5-like fimbria subunit peptides (Lb1(f) peptides) of P5-like fimbria proteins from various Haemophilus influenzae strains. The peptides are used for diagnosis, prevention, and treatment of Haemophilus influenzae infections, such as otitis media, sinusitis, conjunctivitis, or lower respiratory tract infection. The peptides may also be used in vaccines against H. influenzae. Antibodies and probes from the present invention can be used for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and AAZ91201 to AAZ91252, represent sequences used in the exemplification of the present invention

Sequence 19 AA;

Query Match 59.7%; Score 43; DB 3; Length 19;  
Best Local Similarity 77.8%; Pred. No. 1.5;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RSDYKFDN 9  
|||  
DQ 1 RSDYKFDN 9  
|||

RESULT 16  
AAY79965  
ID AAY79965 standard; peptide; 19 AA.  
AC AAY79965;  
XX  
XX 15-MAY-2000 (first entry)  
XX  
XX Non-typeable H. influenzae group 1 Lb1(f) peptide NTHI-484.  
XX  
XX Vaccine; non-typeable Haemophilus influenzae; nTHI; infection;  
XX chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
XX lipoprotein D; Lb1(f); immunogenic; antigenic; otitis media; sinusitis;  
XX conjunctivitis; lower respiratory tract infection.  
XX  
XX Haemophilus influenzae.  
XX  
XX WO9964067-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 28-MAY-1999; 99WO-US011980.  
XX  
XX 11-JUN-1998; 98GB-00012613.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX (OHIS ) UNIV OHIO STATE RES FOUND.  
XX  
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
XX WPI; 2000-116457/10.  
XX  
XX  
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
XX Haemophilus influenzae.  
XX  
XX Example 1; Page 29; 68pp; English.

The present invention describes antigenic P5-like fimbria subunit peptides (Lb1(f) peptides) of P5-like fimbria proteins from various Haemophilus influenzae strains. The peptides are used for diagnosis, prevention, and treatment of Haemophilus influenzae infections, such as otitis media, sinusitis, conjunctivitis, or lower respiratory tract infection. The peptides may also be used in vaccines against H. influenzae. Antibodies and probes from the present invention can be used for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and AAZ91201 to AAZ91252, represent sequences used in the exemplification of the present invention

```

SQ Sequence 19 AA;
  Query Match          59.7%; Score 43; DB 3; Length 19;
  Best Local Similarity 77.8%; Pred. No. 1.5;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDYDN 9
   |||||:::
Db 1 RSDYKFDYED 9

RESULT 17
AAAY79956
ID AAY79956 standard; peptide; 19 AA.
AC AAY79956;
XX
DT 15-MAY-2000 (first entry)
DE Non-typeable H. influenzae group 1 LB1(f) peptide N152NP.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX
PN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
PN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
PT Novel antigenic P5-like fimbrin subunit peptides used in vaccines against
PT Haemophilus influenza.
XX
PS Example 1; Page 29; 68pp; English.
XX
CC The present invention describes antigenic P5-like fimbrin subunit
CC peptides (LB1(f) peptides) of P5-like fimbrin proteins from various
CC Haemophilus influenzae strains. The peptides are used for diagnosis,
CC prevention, and treatment of Haemophilus influenzae infections, such as
CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
CC infection. The peptides may also be used in vaccines against H.
CC influenzae. Antibodies and probes from the present invention can be used
CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 19 AA;
  Query Match          59.7%; Score 43; DB 3; Length 19;
  Best Local Similarity 77.8%; Pred. No. 1.5;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDYDN 9
   |||||:::
Db 1 RSDYKFDYED 9

RESULT 19
AAAY79955
ID AAY79955 standard; peptide; 19 AA.
AC AAY79955;
XX
DT 15-MAY-2000 (first entry)
DE Non-typeable H. influenzae group 1 LB1(f) peptide N1128.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbrin protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX

```

PN WO9964067-A2.  
 PD 16-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 DR WPI; 2000-116457/10.  
 XX  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenza.  
 XX  
 PS Example 1; Page 29; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AAZ91201 to AAZ91252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 19 AA;  
 Query Match 59.7%; Score 43; DB 3; Length 19;  
 Best Local Similarity 77.8%; Pred. No. 1.5;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RSDYKFDYN 9  
 Db |||||:::  
 1 RSDYKFVED 9  
 RESULT 20  
 AAB47439  
 ID AAB47439 standard; peptide; 19 AA.  
 XX  
 AC AAB47439;  
 XX  
 DT 31-OCT-2001 (first entry)  
 XX  
 DE LB1(f) containing peptide from strain ntHi-1128 (Group 1 type).  
 XX  
 KW surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; ntHi; LB1(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO200161013-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 XX 13-FEB-2001; 2001WO-EP001556.  
 XX  
 PR 15-FEB-2000; 2000GB-00003502.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Berthet FJ, Denoel P, Poolman J, Thonnard J;  
 XX  
 DR WPI; 2001-522599/57.  
 XX

PT Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 XX media and conjunctivitis.  
 XX  
 XX Claim 1; Page 26; 29pp; English.  
 PS  
 CC The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (ntHi). Each of these  
 CC peptides contain an LB1(f) peptide which is a 19 amino acid peptide  
 CC derived from the sequence of MOMP P5 from strain ntHi1128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX  
 SQ Sequence 19 AA;  
 Query Match 59.7%; Score 43; DB 4; Length 19;  
 Best Local Similarity 77.8%; Pred. No. 1.5;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RSDYKFDYN 9  
 Db |||||:::  
 1 RSDYKFVED 9  
 RESULT 21  
 AAB20881  
 ID AAB20881 standard; peptide; 20 AA.  
 XX  
 AC AAB20881;  
 XX  
 DT 03-JAN-2001 (first entry)  
 XX  
 DE LB1gr1 peptide SEQ ID NO:19.  
 XX  
 KW Immunoglobulin E; IgE; immunogenic; immunogen; Protein D; carrier;  
 KW prostate cancer; Haemophilus influenzae; vaccine; infectious disease;  
 KW malaria; cytostatic; antiallergic; nontropic; neuroprotective;  
 KW protozoacide; Alzheimer's disease; allergy.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1 /note= "acetylated"  
 FT Modified-site 20 /note= "amidated"  
 FT  
 XX WO2000050077-A1.  
 PN  
 XX 31-AUG-2000.  
 XX  
 XX 22-FEB-2000; 2000WO-EP001457.  
 XX  
 PR 25-FEB-1999; 99GB-00004405.  
 PR 25-FEB-1999; 99GB-00004408.  
 PR 25-FEB-1999; 99GB-00004412.  
 PR 13-AUG-1999; 99GB-00019260.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Coste M, Lobet Y, Van-Mechelen MP, Verriest C;  
 XX  
 DR WPI; 2000-572040/53.  
 XX  
 PT Immunogens and vaccine comprising the immunogen useful for preventing and  
 PT treating infectious diseases e.g. malaria and chronic disease e.g.

PT cancer, comprises peptide and carrier from protein D of influenzae.  
 XX Example 14; Page 34; 53pp; English.  
 XX  
 CC The present invention describes an immunogen (I) comprising a peptide  
 CC (Ia) and a carrier (Ib) derived from protein D of Haemophilus influenzae  
 CC or its fragment. Also described are: (1) a vaccine comprising (I), and an  
 CC excipient; (2) preparation of (I), comprising conjugating a peptide to  
 CC protein D or its fragment; and (3) preparation of a vaccine of (1),  
 CC comprising formulating (I) with an excipient. (I) has cytostatic,  
 CC antiallergic, nootropic, neuroprotective and protozoacide activities. (I)  
 CC and the vaccine are useful for the manufacture of a medicament for  
 CC preventing and treating infectious diseases such as malaria or chronic  
 CC disease such as cancer, Alzheimer's disease or allergy in a patient.  
 CC Unlike prior art immunogens, (I) induces high levels of antipeptide  
 CC immune responses while inducing a moderate humoral response against the  
 CC carrier. The present sequence represents an LBIgri peptide which was  
 CC coupled through an additional C-terminal cysteine via maleimide to  
 CC protein D in an example from the present invention  
 XX  
 SQ Sequence 20 AA;  
 Query Match 59.7%; Score 43; DB 3; Length 20;  
 Best Local Similarity 77.8%; Pred. No. 1.6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKFYDN 9  
 |||||:  
 Db 1 RSDYKFYED 9  
 RESULT 22  
 AAB47443  
 ID AAB47443 standard; peptide; 28 AA.  
 XX  
 AC AAB47443;  
 XX  
 DT 31-OCT-2001 (first entry)  
 XX  
 DE Entire 3rd loop from strain nHi-1128 (Group 1 type).  
 XX  
 KW surface exposed loop; major outer membrane protein P5; MOMP P5;  
 KW non-typeable H. influenzae; nHi; LBI(f) peptide; B cell epitope;  
 KW otitis media; sinusitis; conjunctivitis;  
 KW lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO200161013-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 13-FEB-2001; 2001WO-EP001556.  
 XX  
 PR 15-FEB-2000; 2000GB-00003502.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Berthet FJ, Denoel P, Poolman J, Thonnard J;  
 XX  
 DR WPI; 2001-522599/57.  
 XX  
 PT Recombinant bacterial outer membrane protein where one or more surface-  
 PT exposed loops are modified is useful as a vaccine to prevent or treat  
 PT Haemophilus influenzae infection or associated disease, e.g., otitis  
 PT media and conjunctivitis.  
 XX  
 PS Claim 2; Page 26; 29pp; English.  
 XX  
 CC The sequences given in AAB47439-46 represent peptides which may be used  
 CC to replace one or more surface exposed loops of major outer membrane  
 CC protein P5 (MOMP P5) of non-typeable H. influenzae (nHi). Each of these  
 CC peptides contain an LBI(f) peptide which is a 19 amino acid peptide

CC derived from the sequence of MOMP P5 from strain nHill128, representing  
 CC amino acids Arg117 to Gly135. This peptide represents the third exposed  
 CC loop of P5 and is a potential B cell epitope. The loops of the invention  
 CC are modified in terms of being in a non-native environment in the  
 CC recombinant outer membrane protein. The modified MOMP P5 may be used to  
 CC induce an immune response in a mammal to prevent or treat Haemophilus  
 CC influenzae infection or associated disease, e.g., otitis media,  
 CC sinusitis, conjunctivitis, or lower respiratory tract infection  
 XX  
 SQ Sequence 28 AA;  
 Query Match 59.7%; Score 43; DB 4; Length 28;  
 Best Local Similarity 77.8%; Pred. No. 2.3;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKFYDN 9  
 |||||:  
 Db 3 RSDYKFYED 11  
 RESULT 23  
 AAW67581  
 ID AAW67581 standard; peptide; 40 AA.  
 XX  
 AC AAW67581;  
 XX  
 DT 02-MAR-1999 (first entry)  
 XX  
 DE Synthetic chimera fimbria/T-cell epitope peptide LBI.  
 XX  
 KW Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope;  
 KW immunogenic composition; immune response.  
 XX  
 OS Synthetic.  
 XX  
 PN US5843464-A.  
 XX  
 PD 01-DEC-1998.  
 XX  
 PF 02-JUN-1995; 95US-00460502.  
 XX  
 PR 02-JUN-1995; 95US-00460502.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Kaumaya PTP, Bakaletz LO;  
 XX  
 DR WPI; 1999-044514/04.  
 XX  
 PT Synthetic chimeric fimbria peptide - useful for vaccination against non-  
 PT typable Haemophilus influenzae.  
 XX  
 PS Claim 4; Col 4; 16pp; English.  
 XX  
 CC The invention relates to the manufacture of a synthetic chimeric peptide  
 CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via  
 CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is  
 CC used in immunogenic compositions which induce an immune response against  
 CC non-typable Haemophilus influenzae. This sequence represents an example  
 CC of a chimeric fimbria/T-cell epitope peptide and is designated LBI. The  
 CC peptide comprises a 19 amino acid sequence corresponding to amino acids  
 CC 117-135 of the fimbria protein, the linker sequence and amino acid 288-  
 CC 302 of the measles virus fusion protein (a T-cell epitope)  
 XX  
 SQ Sequence 40 AA;  
 Query Match 59.7%; Score 43; DB 2; Length 40;  
 Best Local Similarity 77.8%; Pred. No. 3.4;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKFYDN 9  
 |||||:  
 Db 1 RSDYKFYED 9



RESULT 24  
 AAY79986  
 ID AAY79986 standard; peptide; 40 AA.  
 AC AAY79986;  
 XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Measles virus fusion protein T-cell promiscuous epitope.  
 XX  
 KW vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; Lbl(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Measles virus.  
 OS Synthetic.  
 XX  
 FN W09964067-A2.  
 XX  
 XX 16-DEC-1999.  
 XX  
 XX 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 PA (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 DR WPI; 2000-116457/10.  
 XX  
 PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenzae.  
 XX  
 PS Example 4; Page 38; 68pp; English.  
 XX  
 CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (Lb1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 40 AA;  
 Query Match 59.7%; Score 43; DB 3; Length 40;  
 Best Local Similarity 77.8%; Pred. No. 3.4;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 RSDYKFDYDN 9  
 DB 1 RSDYKFDYDN 9  
 |||||:  
 RESULT 25  
 ADA25172  
 ID ADA25172 standard; peptide; 40 AA.  
 XX  
 AC ADA25172;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Chimeric fimbria peptide Lb1.  
 XX  
 KW fimbria; non-typeable Haemophilus influenzae; NTHi infection;

otitis media.  
 KW  
 XX Chimeric.  
 OS  
 OS Synthetic.  
 OS Haemophilus influenzae.  
 OS Measles virus.  
 XX  
 XX US6436405-B1.  
 PN  
 XX 20-AUG-2002.  
 PD  
 XX 04-SEP-1998; 98US-00148711.  
 PF  
 XX 02-JUN-1995; 95US-00460502.  
 PR  
 XX (OHIS ) UNIV OHIO STATE.  
 PA  
 XX Bakaletz LO, Kaumaya PTP;  
 PI  
 XX WPI; 2003-615247/58.  
 DR  
 XX Synthetic chimeric fimbria peptide, useful for treating Haemophilus  
 XX influenzae infections.  
 PT  
 PT Claim 10; Col 4; 16pp; English.  
 PS  
 XX The invention relates to a synthetic chimeric fimbria peptide. The  
 CC peptide is useful for treating a non-typeable Haemophilus influenzae  
 CC (NTHi) infection and otitis media. The synthetic peptides do not require  
 CC tedious purification techniques. The present sequence represents the  
 CC amino acid sequence of the chimeric fimbria peptide Lb1.  
 XX  
 XX Sequence 40 AA;  
 SQ  
 Query Match 59.7%; Score 43; DB 6; Length 40;  
 Best Local Similarity 77.8%; Pred. No. 3.4;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 RSDYKFDYDN 9  
 DB 1 RSDYKFDYDN 9  
 |||||:  
 RESULT 26  
 ADC89661  
 ID ADC89661 standard; peptide; 40 AA.  
 XX  
 AC ADC89661;  
 AC  
 XX 01-JAN-2004 (first entry)  
 DT  
 XX H. influenzae fimbria peptide/T cell epitope chimaera Lb1.  
 DE  
 XX Fimbria; T cell epitope; vaccine; otitis media; auditory;  
 KW anti-inflammatory; Lb1.  
 KW  
 XX Chimeric.  
 OS  
 OS Haemophilus influenzae.  
 OS Measles virus.  
 XX  
 XX US2003113344-A1.  
 PN  
 XX 19-JUN-2003.  
 PD  
 XX 19-AUG-2002; 2002US-00223711.  
 PF  
 XX 04-SEP-1998; 98US-00148711.  
 PR  
 XX (BAKA/) BAKALETZ L O.  
 PA (KAUM/) KAUMAYA P T P.  
 PA  
 PI Bakaletz LO, Kaumaya PTP;  
 XX

DR WPI; 2003-810881/76.

XX Novel synthetic chimeric fimbrin peptide LB1 or LB2 comprising a first

PT peptide unit, T cell epitope as second peptide unit and third linker

PT peptide unit, useful for preventing or reducing severity of otitis media.

XX

PS Claim 8; SEQ ID NO 10; 15pp; English.

XX

CC The invention relates to a synthetic chimaeric fimbrin peptide LB1 or LB2

CC comprises a first peptide unit derived from H. influenzae fimbrin, a

CC second peptide unit containing a T cell epitope and a third linker

CC peptide which connects the first peptide to the second. The chimaeric

CC peptide is useful for inducing an immune response in animals against non-

CC typable Haemophilus influenzae (NTHi) and for preventing or reducing the

CC adherence of NTHi to host cells thereby preventing or reducing the

CC severity of otitis media. The present sequence is an H. influenzae

CC fimbrin peptide/measles virus T cell epitope chimaeric peptide of the

CC invention, LB1.

XX

XX Sequence 40 AA;

XX

Query Match 59.7%; Score 43; DB 7; Length 40;

Best Local Similarity 77.8%; Pred. No. 3.4; Indels 0; Gaps 0;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9

Db 1 RSDYKFYED 9

|||||::

1 RSDYKFYED 9

138 RSDYKFYED 146

RESULT 27

AAR66294

ID AAR66294 standard; protein; 359 AA.

XX

AC AAR66294;

XX

DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 09-AUG-1995 (first entry)

XX

DE Non-typable Haemophilus influenza (NTHi) fimbrin protein.

XX

XX Fimbrin protein; vaccine; otitis media.

XX

XX Haemophilus influenzae.

OS

OS

FH Key Location/Qualifiers

FT Region 22..33

FT Peptide /label= amino terminus

FT 234..249

FT /label= internal CNBr fragment

XX

XX WO9426304-A1.

XX

XX 24-NOV-1994.

XX

XX 12-MAY-1994; 94WO-US005477.

XX

XX 18-MAY-1993; 93US-00065442.

XX

XX (OHIS ) UNIV OHIO STATE RES FOUND.

XX

XX Kolattukudy PE, Bakaletz LO, Sirakova T;

XX

XX WPI; 1995-006359/01.

DR N-PSDB; AAQ78916.

XX

XX Vaccine comprising non-typable Haemophilus influenza fimbrin protein -

PT useful in studying, preventing or reducing the severity of otitis media,

PT also fimbrin protein and DNA.

XX

XX Disclosure; Fig 5; 45pp; English.

XX

CC The fimbrin proteins from 15 randomly selected type b and non- typable

CC clinical isolates of Haemophilus influenzae share common epitopes. Thus

CC fimbrin isolated from non-typable Haemophilus influenzae 1128 strain is

CC a particularly suitable immunogen to protect against the different non-

CC typable HJ. influenzae that cause otitis media. Fimbrin protein is

CC produced by culturing a transformed microbial host, pref. E.coli,

CC Sporodoptera frugiperda or a mucosal pathogen. Fimbrin migrates in

CC polyacrylamide gels to a posn. equiv. to a mol. wt. of 25.5 kD or 37.5

CC kD. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003

CC to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)

XX

XX Sequence 359 AA;

XX

Query Match 59.7%; Score 43; DB 2; Length 359;

Best Local Similarity 77.8%; Pred. No. 39;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFYDN 9

Db 138 RSDYKFYED 146

|||||::

1 RSDYKFYED 9

138 RSDYKFYED 146

RESULT 28

ABU22435

ID ABU22435 standard; protein; 696 AA.

XX

AC ABU22435;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #7962.

XX

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

XX Burkholderia mallei.

OS

OS

PN WO200277183-A2.

XX

XX 03-OCT-2002.

PD

XX 21-MAR-2002; 2002WO-US009107.

PF

XX 21-MAR-2001; 2001US-00815242.

PR

XX 06-SEP-2001; 2001US-00948993.

PR

XX 25-OCT-2001; 2001US-0342923P.

PR

XX 08-FEB-2002; 2002US-00072851.

PR

XX 06-MAR-2002; 2002US-0362699P.

PR

XX (ELIT-) ELITRA PHARM INC.

PA

XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

XX WPI; 2003-029926/02.

DR

XX N-PSDB; ACA26305.

DR

XX

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

XX Claim 25; SEQ ID NO 50359; 1766pp; English.

XX

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 696 AA;

Query Match 59.7%; Score 43; DB 6; Length 696;  
 Best Local Similarity 46.2%; Pred. No. 81;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RSDYKFDNKRID 13  
 ||||| : : : :  
 Db 133 RSDYRIFQNRSD 145

RESULT 29  
 AAY79959  
 ID AAY79959 standard; peptide; 19 AA.  
 XX  
 AC AAY79959;  
 XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Non-typeable H. influenzae group 1 LB1(f) peptide N10567RM.  
 XX  
 KW Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO9964067-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US011980.  
 XX  
 PR 11-JUN-1998; 98GB-00012613.  
 XX  
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA (OHIS) UNIV OHIO STATE RES FOUND.  
 XX  
 PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX  
 DR WPI; 2000-116457/10.  
 XX  
 PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 PT Haemophilus influenzae.  
 XX  
 PS Example 1; Page 29; 68pp; English.

CC The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,

CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 19 AA;

Query Match 58.3%; Score 42; DB 3; Length 19;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDYD 8  
 ||||| :  
 Db 1 RSDYKFDYE 8

RESULT 30  
 AAY79970  
 ID AAY79970 standard; peptide; 19 AA.  
 XX  
 AC AAY79970;

XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Non-typeable H. influenzae group 1 LB1(f) peptide N1657MBE.

XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;  
 KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;  
 KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;  
 KW conjunctivitis; lower respiratory tract infection.

XX Haemophilus influenzae.

XX WO9964067-A2.

XX 16-DEC-1999.

XX 28-MAY-1999; 99WO-US011980.

XX 11-JUN-1998; 98GB-00012613.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX (OHIS) UNIV OHIO STATE RES FOUND.

XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;

XX WPI; 2000-116457/10.

XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenzae.

XX Example 1; Page 29; 68pp; English.

XX The present invention describes antigenic P5-like fimbria subunit  
 CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various  
 CC Haemophilus influenzae strains. The peptides are used for diagnosis,  
 CC prevention, and treatment of Haemophilus influenzae infections, such as  
 CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 CC infection. The peptides may also be used in vaccines against H.  
 CC influenzae. Antibodies and probes from the present invention can be used  
 CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 CC AA291201 to AA291252, represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 19 AA;

Query Match 58.3%; Score 42; DB 3; Length 19;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 RSDYKFFD 8
Db 1 RSDYKFYE 8

RESULT 31
AA79992
ID AA79992 standard; peptide; 19 AA.
XX
AC AA79992;
XX
DT 15-MAY-2000 (first entry)
XX
DE Non-typeable H. influenzae protien P5 Group 1 type peptide.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nHi; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX
FN WO9964067-A2.
XX
PD 16-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US011980.
XX
PR 11-JUN-1998; 98GB-00012613.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
DR WPI; 2000-116457/10.
XX
PN Novel antigenic P5-like fimbria subunit peptides used in vaccines against
PD Haemophilus influenza.
XX
PF Example 1; Page 29; 68pp; English.
XX
PR The present invention describes antigenic P5-like fimbria subunit
PA peptides (LB1(f) peptides) of P5-like fimbria proteins from various
PA Haemophilus influenza strains. The peptides are used for diagnosis,
PA prevention, and treatment of Haemophilus influenza infections, such as
XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX infection. The peptides may also be used in vaccines against H.
XX influenza. Antibodies and probes from the present invention can be used
XX for diagnosis of H. influenza infection. AA79955 to AA79993, and
XX AA291201 to AA291252, represent sequences used in the exemplification of
XX the present invention
XX
SQ Sequence 19 AA;

Query Match 58.3%; Score 42; DB 3; Length 19;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RSDYKFFD 8
Db 1 RSDYKFYE 8

RESULT 33
AA79960
ID AA79960 standard; peptide; 19 AA.
XX
AC AA79960;
XX
DT 15-MAY-2000 (first entry)
XX
DE Non-typeable H. influenzae group 1 LB1(f) peptide N86027NP.
XX
KW Vaccine; non-typeable Haemophilus influenzae; nHi; infection;
KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
KW conjunctivitis; lower respiratory tract infection.
XX
OS Haemophilus influenzae.
XX
FN WO9964067-A2.
XX

```

```

PD 16-DEC-1999.
XX
XX 28-MAY-1999; 99WO-US011980.
XX
XX 11-JUN-1998; 98GB-00012613.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX WPI; 2000-116457/10.
XX
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.
XX
XX Example 1; Page 29; 68pp; English.
XX
XX The present invention describes antigenic P5-like fimbria subunit
XX peptides (LBI(f) peptides) of P5-like fimbria proteins from various
XX Haemophilus influenzae strains. The peptides are used for diagnosis,
XX prevention, and treatment of Haemophilus influenzae infections, such as
XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX infection. The peptides may also be used in vaccines against H.
XX influenzae. Antibodies and probes from the present invention can be used
XX for diagnosis of H. influenzae infection. AAY7955 to AAY7993, and
XX AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX the present invention
XX
XX Sequence 19 AA;
XX
XX Query Match 58.3%; Score 42; DB 3; Length 19;
XX Best Local Similarity 87.5%; Pred. No. 2.2;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 RSDYKFPD 8
XX Db 1 RSDYKFPY 8
XX
XX RESULT 35
XX AAY79962
XX ID AAY79962 standard; peptide; 19 AA.
XX
XX AC AAY79962;
XX
XX DT 15-MAY-2000 (first entry)
XX
XX DE Non-typeable H. influenzae group 1 LBI(f) peptide N166NP.
XX
XX KW Vaccine; non-typeable Haemophilus influenzae; nThi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX
XX OS Haemophilus influenzae.
XX
XX PN WO9964067-A2.
XX
XX PD 16-DEC-1999.
XX
XX PF 28-MAY-1999; 99WO-US011980.
XX
XX PR 11-JUN-1998; 98GB-00012613.
XX
XX PR (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
XX DR WPI; 2000-116457/10.
XX
XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.
XX
XX Example 1; Page 29; 68pp; English.
XX
XX The present invention describes antigenic P5-like fimbria subunit
XX peptides (LBI(f) peptides) of P5-like fimbria proteins from various
XX Haemophilus influenzae strains. The peptides are used for diagnosis,
XX prevention, and treatment of Haemophilus influenzae infections, such as
XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX infection. The peptides may also be used in vaccines against H.
XX influenzae. Antibodies and probes from the present invention can be used
XX for diagnosis of H. influenzae infection. AAY7955 to AAY7993, and
XX AAZ91201 to AAZ91252, represent sequences used in the exemplification of
XX the present invention
XX
XX Sequence 19 AA;
XX
XX Query Match 58.3%; Score 42; DB 3; Length 19;
XX Best Local Similarity 87.5%; Pred. No. 2.2;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 RSDYKFPD 8
XX Db 1 RSDYKFPY 8
XX
XX RESULT 34
XX AAY79961
XX ID AAY79961 standard; peptide; 19 AA.
XX
XX AC AAY79961;
XX
XX DT 15-MAY-2000 (first entry)
XX
XX DE Non-typeable H. influenzae group 1 LBI(f) peptide NTHI-476.
XX
XX KW Vaccine; non-typeable Haemophilus influenzae; nThi; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LBI(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX
XX OS Haemophilus influenzae.
XX
XX PN WO9964067-A2.
XX
XX PD 16-DEC-1999.
XX
XX PF 28-MAY-1999; 99WO-US011980.
XX
XX PR 11-JUN-1998; 98GB-00012613.
XX
XX PR (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX
XX PI Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX
XX DR WPI; 2000-116457/10.
XX
XX PT Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX Haemophilus influenza.

```

```

XX SQ Sequence 19 AA;
    Query Match      58.3%; Score 42; DB 3; Length 19;
    Best Local Similarity 77.8%; Pred. No. 2.2;
    Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDYD 9
Db 1 RSDYKFDYD 9

RESULT 36
AA79987
ID AAY79987 standard; peptide; 19 AA.
XX AC
XX XX
XX XX
DT 15-MAY-2000 (first entry)
XX AC
XX AAY79987;
XX DT
XX 15-MAY-2000 (first entry)
XX DE
XX Non-typeable H. influenzae 10567RM Group 1 type peptide.
XX DE
XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS
XX Haemophilus influenzae.
XX PN WO9964067-A2.
XX XX
XX 16-DEC-1999.
XX PD
XX 28-MAY-1999; 99WO-US011980.
XX PF
XX 11-JUN-1998; 98GB-00012613.
XX PR
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA (OHIS ) UNIV OHIO STATE RES FOUND.
XX XX
XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;
XX PI
XX WPI; 2000-116457/10.
XX DR
XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against
XX PT Haemophilus influenza.
XX PT
XX Example 1; Page 29; 68pp; English.
XX PS
XX The present invention describes antigenic P5-like fimbria subunit
XX CC peptides (LB1(f) peptides) of P5-like fimbria proteins from various
XX CC Haemophilus influenzae strains. The peptides are used for diagnosis,
XX CC prevention, and treatment of Haemophilus influenzae infections, such as
XX CC otitis media, sinusitis, conjunctivitis, or lower respiratory tract
XX CC infection. The peptides may also be used in vaccines against H.
XX CC influenzae. Antibodies and probes from the present invention can be used
XX CC for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and
XX CC AA291201 to AA291252, represent sequences used in the exemplification of
XX CC the present invention
XX XX
XX SQ Sequence 19 AA;
    Query Match      58.3%; Score 42; DB 3; Length 19;
    Best Local Similarity 87.5%; Pred. No. 2.2;
    Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDYKFDYD 8
Db 1 RSDYKFDYD 8

RESULT 37
AA79971
ID AAY79971 standard; peptide; 19 AA.
XX AC
XX AAY79971;
XX XX
XX 15-MAY-2000 (first entry)
XX DT
XX Non-typeable H. influenzae group 1 LB1(f) peptide N214NP.
XX DE
XX Vaccine; non-typeable Haemophilus influenzae; nH1; infection;
XX KW chimeric protein; Haemophilus influenzae; P5-like fimbria protein;
XX KW lipoprotein D; LB1(f); immunogenic; antigenic; otitis media; sinusitis;
XX KW conjunctivitis; lower respiratory tract infection.
XX OS
XX Haemophilus influenzae.

```

XX WO9964067-A2.  
 XX 16-DEC-1999.  
 XX 28-MAY-1999; 99WO-US011980.  
 XX 11-JUN-1998; 98GB-00012613.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Bakaletz LO, Cohen J, Dequesne G, Lobet Y;  
 XX WPI; 2000-116457/10.  
 XX Novel antigenic P5-like fimbria subunit peptides used in vaccines against  
 XX Haemophilus influenza.  
 XX Example 1; Page 29; 68pp; English.  
 XX The present invention describes antigenic P5-like fimbria subunit  
 XX peptides (LBI(f) peptides) of P5-like fimbria proteins from various  
 XX Haemophilus influenzae strains. The peptides are used for diagnosis,  
 XX prevention, and treatment of Haemophilus influenzae infections, such as  
 XX otitis media, sinusitis, conjunctivitis, or lower respiratory tract  
 XX infection. The peptides may also be used in vaccines against H.  
 XX influenzae. Antibodies and probes from the present invention can be used  
 XX for diagnosis of H. influenzae infection. AAY79955 to AAY79993, and  
 XX AA291201 to AA291252, represent sequences used in the exemplification of  
 XX the present invention  
 XX Sequence 19 AA;  
 SQ Sequence 19 AA;  
 Query Match 58.3%; Score 42; DB 3; Length 19;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RSDYKFYD 8  
 |||||:  
 Db 1 RSDYKFYE 8  
 |||||:  
 RESULT 39  
 ADB11406  
 ID ADB11406 standard; protein; 329 AA.  
 XX ADB11406;  
 XX 20-NOV-2003 (first entry)  
 XX Allotiococcus otitis antigenic protein SEQ ID NO:5654.  
 XX Allotiococcus otitidis; antigenic protein; immunogenic; immunisation;  
 XX gene therapy; Gram-positive bacterium; infection.  
 XX Allotiococcus otitis.  
 XX WO2003048304-A2.  
 XX 12-JUN-2003.  
 XX 25-NOV-2002; 2002WO-US036123.  
 XX 29-NOV-2001; 2001US-0333777P.  
 XX 18-NOV-2002; 2002US-0426742P.  
 XX (AMHP ) WYETH HOLDINGS CORP.  
 XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
 XX WPI; 2003-505284/47.  
 XX N-PSDB; ADB11409.

XX New Allotiococcus otitidis polynucleotides and polypeptides, useful for  
 XX treating and diagnosing diseases, drug screening assays and monitoring of  
 XX effects during drug clinical trials.  
 XX Claim 33; SEQ ID NO 5654; 1019pp; English.  
 XX The present invention describes an isolated polynucleotide (I) of  
 XX Allotiococcus otitidis genomic DNA, which encodes an antigenic protein.  
 XX Allotiococcus otitidis is a Gram-positive bacterium. Also described: (1)  
 XX an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
 XX expression vector comprising the novel isolated polynucleotide (I); its  
 XX complement, degenerate variant or fragment; (3) a genetically engineered  
 XX host cell, transfected, transformed or infected with the vector of (2);  
 XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 XX composition comprising the polypeptide, its complement, biological  
 XX equivalent or fragment, or the polynucleotide that is comprised in the  
 XX expression vector; (6) a pharmaceutical composition comprising the  
 XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 XX of the polypeptides of (1), their biological equivalent or fragment; (8)  
 XX immunising against Allotiococcus otitidis by administering to a host the  
 XX immunogenic composition; (9) detecting and/or identifying Allotiococcus  
 XX otitidis in the biological sample; (10) a kit comprising a container  
 XX containing the novel polynucleotide, its degenerate variant or fragment,  
 XX or the antibody of (4); and (11) producing a polypeptide by culturing the  
 XX genetically engineered host cell under conditions suitable to produce the  
 XX polypeptide from the culture. (I) can be used in gene therapy. The  
 XX polynucleotides, polypeptides, antibodies and compositions of the present  
 XX invention can be used for treating and diagnosing diseases, drug  
 XX screening assays and monitoring of effects during drug clinical trials.  
 XX The polynucleotides are useful for expressing and detecting Allotiococcus  
 XX otitidis. The present sequence represents an Allotiococcus otitidis  
 XX antigen protein from the present invention.  
 SQ Sequence 329 AA;  
 Query Match 58.3%; Score 42; DB 6; Length 329;  
 Best Local Similarity 58.3%; Pred. No. 53;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 SDYKFDNKRID 13  
 ||:|:|:|:|:  
 Db 127 SDFKQDRSLD 138  
 ||:|:|:|:|:  
 RESULT 40  
 AAR85450  
 ID AAR85450 standard; protein; 338 AA.  
 XX AAR85450;  
 XX 15-FEB-1996 (first entry)  
 XX Nontypable H. influenzae P5 protein.  
 XX P5 outer membrane protein; vaccine; otitis media; sinusitis;  
 XX chronic pulmonary obstructive disease.  
 XX Haemophilus influenzae.  
 XX OS  
 XX Location/Qualifiers  
 FH Key  
 FT Misc-difference 195  
 /note= "amino acid at position 195 is not identified in  
 the specification"  
 FT Misc-difference 311  
 /note= "amino acid at position 311 is not identified in  
 the specification"  
 FT  
 XX EP680765-A1.  
 XX 08-NOV-1995.  
 XX 02-MAY-1995; 95EP-00302996.

XX 05-MAY-1994; 94US-00210394.  
XX  
XX PA (AMCY ) AMERICAN CYANAMID CO.  
XX PI Zlotnick GW;  
XX  
XX DR WPI; 1995-375029/49.  
XX  
XX PT Purified H. influenzae P5 outer membrane protein - used for preventing  
XX PT reducing susceptibility to or treating H. influenzae infections.  
XX PS Disclosure; Page 7-8; 16pp; English.  
XX  
XX CC Nontypable H. influenzae HI outer membrane protein P5 was isolated by  
XX CC extraction of the outer membrane with detergents and cation-exchange  
XX CC chromatography. P5 (or its peptide fragments) are used in vaccines for  
XX CC prevention of H. influenzae infections implicated in otitis media,  
XX CC sinusitis and chronic pulmonary obstructive disease  
XX SQ Sequence 338 AA;  
Query Match 58.3%; Score 42; DB 2; Length 338;  
Best Local Similarity 87.5%; Fred. No. 54;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RSDYKFD 8  
Db 119 RSDYKFYE 126  
|||||:  
Search completed: November 24, 2004, 09:21:43  
Job time : 57.663 secs